

Anwendbarkeit von Zufallsstichproben bei der Abschlußprüfung

Inaugural-Dissertation
zur Erlangung der Doktorwürde
einer Hohen Staatswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität
zu München

Vorgelegt von
Karl Elmendorff
aus Düsseldorf

1962

Anwendbarkeit von Zufallsstichproben bei Abschlußprüfungen

Anwendbarkeit von Zufallsstichproben bei der Abschlußprüfung

Inaugural-Dissertation
zur Erlangung der Doktorwürde
einer Hohen Staatswirtschaftlichen Fakultät
der Ludwig-Maximilians-Universität
zu München



Vorgelegt von
Karl Elmendorff
aus Düsseldorf
1962

Referent: Prof. Dr. H. Kellerer

Korreferent: Prof. Dr. O. Hintner

Tag der mündlichen Prüfung: 18. 12. 1962

INHALT

Seite

Einleitung

I. Ziel der Untersuchung	1
II. Abriß über die Jahresabschlußprüfung, insbesondere über die Prüfungs- technik	4
1. Grundlagen der Jahresabschlußprüfung	4
2. Prüfungstechnik	6
III. Grundbegriffe der Stichprobentheorie	11

Hauptteil

I. Gliederung der Prüfungsfragen	17
II. Prüfung der Vollständigkeit der Rechnungslegung	18
1. Prüfung der Vollständigkeit der Bestandsaufnahme	18
2. Prüfung der Vollständigkeit der Erfassung der Geschäftsvorfälle	19
III. Prüfung der formellen Ordnungsmäßigkeit der Rechnungslegung	19
1. Prüfung der Einhaltung des Systems der doppelten Buchführung	19
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	19
b) Prüfung einer Grundgesamtheit auf Fehlerfreiheit mittels Entdeckungsstichproben	21
c) Anwendbarkeit auf das Prüfungsgebiet	24
2. Prüfung der Einhaltung der sonstigen formellen Ordnungs- vorschriften	25
A) Prüfung der Einhaltung der gesetzlichen Ordnungsvorschriften	25
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	25

b) Entscheidung über Annahme oder Rückweisung einer Grundgesamtheit mittels Annahmestichproben	26
c) Anwendbarkeit auf das Prüfungsgebiet	34
B) Prüfung der Einhaltung der unternehmensinternen Ordnungsvorschriften	38
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	38
b) Schätzung von Fehleranteilen mittels Schätzstichproben . .	40
c) Anwendbarkeit auf das Prüfungsgebiet	49
3. Bedeutung der Ergebnisse der Prüfung der Sicherheit der Rechnungslegung für die Prüfung der materiellen Richtigkeit der Rechnungslegung	53
IV. Prüfung der materiellen Richtigkeit der Rechnungslegung	56
1. Prüfung der Bestände	56
A) Gesamtwertermittlungen	56
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	56
b) Schätzung von Gesamtwerten mittels Schätzstichproben . .	58
c) Anwendbarkeit auf das Prüfungsgebiet	60
d) Beispiele	63
Vollaufnahme und stichprobenweise Bewertung der Bestände	63
Stichprobenweise Aufnahme und Bewertung der Bestände .	65
Stichprobenweise Prüfung einer permanenten Inventur . .	67
Schrittweise Anpassung eines Stichprobenverfahrens an die Struktur der Grundgesamtheit aufgrund schrittweise verbesserter Informationen	69
B) Bestandsanalysen	73
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	73
b) Schätzung von Anteilen und Mittelwerten mittels Schätzstichproben	74
c) Anwendbarkeit auf das Prüfungsgebiet	74
d) Beispiele	76
Analyse eines Debitorenbestandes nach Umfang und Wert bestimmter Größenklassen der Debitoren und Ermittlung ihres Gesamtwerts	76
Kreditfristanalyse in Verbindung mit einer Gesamtwertschätzung der Debitoren	79

2. Prüfung der Geschäftsvorfälle	80
a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen	80
b) Schätzung von Gesamtwerten, Mittelwerten und Anteilen mittels Schätzstichproben	81
c) Anwendbarkeit auf das Prüfungsgebiet, insbesondere Prüfbarkeit der unter Einsatz selbsttätig arbeitender Geräte erstellten Rech- nungslegung mittels Stichproben	82
3. Festsetzung des Vertrauensbereichs und Sicherheitsgrades für Stichproben bei der Prüfung materieller Tatbestände	87

Schluß

Zusammenfassung der Ergebnisse	91
Literaturverzeichnis	95



Digitized by the Internet Archive
in 2026

EINLEITUNG

I. Ziel der Untersuchung

Bei der Prüfung der Rechnungslegung eines Unternehmens steht der Prüfer immer wieder vor der Frage, ob er sich sein Urteil über den Prüfungsstoff aufgrund einer *lückenlosen Prüfung* oder einer *Teilprüfung* bilden soll. Das gilt insbesondere auch für die Jahresabschlußprüfung eines Unternehmens.

Ein sicheres Kriterium für die Auswahl und den Umfang der Teilprüfungen des Prüfungsstoffs ist bisher aber weder in der Literatur noch in der Praxis zu finden.

In der Literatur wird, sofern die Autoren zu diesem Problem überhaupt Stellung nehmen, einhellig die Meinung vertreten, daß ein Urteil über den Prüfungsstoff nicht möglich sei, wenn die Stichproben mehr oder weniger wahllos gestreut dem gesamten Prüfungsgebiet entnommen würden. Bei dieser Methode bleibe die Entdeckung von Fehlern allzusehr dem Zufall überlassen. Es herrscht vielmehr allgemein die Auffassung, daß der Prüfer den besten Einblick in die Einzelheiten der Buchführungsorganisation erhalte, wenn er gewisse *Teilgebiete des Prüfungsstoffs* herausgreife, diese aber dann möglichst vollständig und genau erfasse. Ferner wird es für notwendig erachtet, die nicht intensiv geprüften Teilgebiete kritisch durchzusehen oder stichprobenweise zu prüfen, weil dies zur Erkenntnis der großen Linien der Rechnungslegung und zur Erfassung etwaiger Besonderheiten notwendig sei¹⁾.

Über Auswahl und Umfang der Stichproben schreiben *Adler/Düring/Schmaltz*²⁾: „In welcher Weise und in welchem Umfang der Abschlußprüfer im einzelnen zu prüfen hat, wie umfangreich die Stichproben sein müssen, welche Sachgebiete er besonders untersuchen muß, um die Ordnungs- und Gesetzmäßigkeit festzustellen, richtet sich im einzelnen nach der Berufsübung, für die sich allmählich Grundsätze ordnungsmäßiger Prüfung herausgebildet haben“.

Die genannten Autoren stellen aber anschließend selbst fest³⁾:

„Was unter Grundsätzen ordnungsmäßiger Abschlußprüfung zu verstehen ist, ist von der Berufsorganisation der Wirtschaftsprüfer noch nicht näher umrissen worden.“ Die Autoren weisen ausdrücklich darauf hin, „daß sich eine feste Übung noch nicht gebildet hat und daß bei der Mannigfaltigkeit der wirtschaftlichen Verhältnisse und der organisatorischen Einrichtungen der Gesellschaften das ver-

¹⁾ *Adler, H./Düring, W./Schmaltz, K.*: Rechnungslegung und Prüfung der Aktiengesellschaft, 3. Aufl., Stuttgart 1957, § 135, Textziffer (Tz.) 122 und Tz. 144; *Hintner, O.*: Praxis der Wirtschaftsprüfung, 3. Aufl., Stuttgart 1949, S. 65; *Minz, W.*: Prüfungsmethoden, Die Wirtschaftsprüfung, Jg. 1960, S. 90; u. v. a.

²⁾ a. a. O., § 135, Tz. 115.

³⁾ a. a. O., § 135, Tz. 116.

antwortungsbewußte Ermessen des Abschlußprüfers und die Besonderheiten des Einzelfalls für den Prüfungsumfang entscheidend bleiben müssen. Als allgemeine Richtlinie kann festgehalten werden, daß insbesondere die Prüfungsmethoden von der Größe der zu prüfenden Gesellschaft weitgehend abhängig sind. Bei kleineren Gesellschaften werden verhältnismäßig zahlreiche Prüfungshandlungen (Stichproben) vorgenommen werden müssen, damit sich der Abschlußprüfer von der Ordnungsmäßigkeit der Buchführung und des Jahresabschlusses überzeugen kann. Bei großen Gesellschaften wird das Schwergewicht der Prüfung mehr auf einer Kontrolle der vorhandenen Prüfungseinrichtungen liegen; denn hier wird der Prüfungsstoff regelmäßig so umfangreich sein, daß er allein durch Stichproben nicht durchleuchtet werden kann.“

Ähnlich allgemein sind auch die Richtlinien, nach denen der amerikanische Prüfer über den Umfang seiner Prüfungsstichproben befindet. Als wesentliche Gesichtspunkte, mit deren Hilfe eine Orientierung bei der Bestimmung des Stichprobenumfangs möglich sei, wurden von *Vance/Neter*¹⁾ folgende zusammengestellt, ohne daß die Autoren aus den Quellen nähere Angaben über die Art und Weise der Berücksichtigung dieser Gesichtspunkte entnehmen konnten:

- a) Qualität des Systems der innerbetrieblichen Kontroll- und Prüfungseinrichtungen und Wirksamkeit seiner Anwendung;
- b) Art der durchzuführenden Prüfung (eine reine Bilanzprüfung erfordert weniger, eine eingehende Revision mehr Stichproben);
- c) Umfang der im Prüfungsbericht anzuführenden Einzelheiten;
- d) Im Laufe der Prüfung erkannte Verhältnisse;
- e) Bedeutung der verschiedenen Prüfungsgebiete;
- f) Besondere Wünsche des geprüften Unternehmens hinsichtlich zusätzlicher Prüfungssicherheit.

Der Prüfer muß also die Verantwortung für die Bestimmung der im Einzelfall erforderlichen Prüfungsverfahren sowie des Umfangs der Stichproben selbst übernehmen. Er kann sich dabei auf seine berufliche Urteilkraft, seine praktischen Erfahrungen und auf die Grundsätze ordnungsmäßiger Abschlußprüfung stützen, soweit sie formuliert und anerkannt sind²⁾. Die Gutachten des Hauptfachausschusses des Instituts der Wirtschaftsprüfer geben zwar die Berufsauffassung zu zahlreichen Buchführungs-, Bilanzierungs- und auch Prüfungsfragen wieder; mit dem hier behandelten Spezialgebiet befassen sie sich jedoch bis auf zwei Ausnahmen, die aber für dieses Thema ebenfalls keine wesentlichen Erkenntnisse bringen, nicht. Diese Ausnahmen sind das Fachgutachten 6/34 über die „Heranziehung vorhandener Prüfungseinrichtungen des geprüften Unternehmens durch den Wirtschaftsprüfer³⁾“ und das Fachgutachten 1/37 über „Pflichtprüfung und Unterschlagungsprüfung⁴⁾“.

1) *Vance, L. L./Neter, J.*: Statistical Sampling for Auditors and Accountants, New York 1956, S. 4.

2) Vgl. dazu: Die Fachgutachten und Stellungnahmen des Instituts der Wirtschaftsprüfer auf dem Gebiete der Rechnungslegung und Prüfung, herausgegeben vom Institut der Wirtschaftsprüfer, bearbeitet von *R. Kenntemich*, Düsseldorf 1956, Einleitung S. 6 f.

3) Die Fachgutachten und Stellungnahmen ..., a. a. O., S. 93 f.

4) Die Fachgutachten und Stellungnahmen ..., a. a. O., S. 112 f.

Die Nachteile der bisher üblichen Verfahren fassen Vance/Neter¹⁾ wie folgt zusammen:

- a) Es gibt in der Praxis bisher keine eindeutige Definition der „Qualität der Rechnungslegung“ (accounting quality), die eine klare Bestimmung des angemessenen Umfangs der Stichprobe und der richtigen Deutung der Stichprobenergebnisse gestattet.
- b) Es lassen sich keine objektiven Berechnungen der Risiken von Fehlurteilen und -schlüssen aufgrund der Stichproben anstellen.
- c) Der Prüfer ist bei der Festlegung von Auswahl und Umfang der Stichprobe und bei seiner Urteilsbildung aufgrund des Stichprobenergebnisses allen möglichen subjektiven Einflüssen (z. B. Gesundheitszustand des Prüfers oder Drängen des Unternehmens auf Beendigung der Prüfungsarbeiten) ausgesetzt.

Diese Nachteile bedingen mangelnde Sicherheit im Urteil über die nicht geprüften Teile der Rechnungslegung und führen daher leicht zu willkürlichen, entweder unangemessen geringen oder unwirtschaftlich großen Stichprobenumfängen.

Als festes Fundament für seine Entscheidungen müßte der Prüfer einen Maßstab haben, der ihm ein von subjektiven Einflüssen freies Urteil über das notwendige Ausmaß der Prüfung nach Art, Umfang und Sicherheit ermöglicht²⁾. In dieser Arbeit soll untersucht werden, ob sich mit Hilfe der *Stichprobentheorie* solche Maßstäbe entwickeln lassen.

Der erste Versuch, statistische Theorien im Gebiet der Revision und der Prüfung der Rechnungslegung anzuwenden, galt dem Ziel, die Chancen der Entdeckung von Unterschlagungen und anderen Unregelmäßigkeiten zu berechnen.

In Deutschland geschah dies bereits 1911 mit den Arbeiten von *Schmalenbach* „Über Einrichtungen gegen Unterschlagungen und über Unterschlagungsrevision³⁾“ und insbesondere von *Klein* „Über die Wahrscheinlichkeit der Entdeckung von Fehlern bei Revisionen⁴⁾“. In Amerika behandelte *L. A. Carman* dieses Thema 1933 in seinem Artikel „The Efficacy of Tests“⁵⁾.

Die Überlegungen, die zunächst nur für dieses Teilgebiet angestellt wurden, sind nicht weiter vorangetrieben worden. In jüngster Zeit sind die Bestrebungen wieder aufgenommen worden, die bei der Prüfung der Rechnungslegung auftauchenden Stichprobenprobleme genauer zu untersuchen⁶⁾.

In der in Amerika im Gang befindlichen und, wie die vielfältige Literatur beweist, lebhaften Diskussion stellen die Ausführungen von Vance/Neter einen gewissen Höhepunkt dar. Ihr Buch soll „Revisoren und Wirtschaftsprüfern sowie Studie-

1) a. a. O., S. 7.

2) So auch *Rackles, R.*: Das Problem der repräsentativen Auswahl bei der aktienrechtlichen Abschlußprüfung, Frankfurt/M. 1961, S. 24.

3) Zeitschrift für handelswissenschaftliche Forschung (ZfhF), Jahrg. 1911/12, S. 321 ff.

4) ZfhF, Jahrg. 1911/12, S. 580 ff.

5) The American Accountant, Jahrg. 1933, S. 360 ff.

6) So z. B. *Loitlsberger, E.*: Treuhand- und Revisionswesen, Stuttgart 1961, S. 73 ff.; *Buchner, R.*, Bewußte Auswahl oder Zufallsauswahl bei der aktienrechtlichen Pflichtprüfung, Die Wirtschaftsprüfung, Dezember 1961, Jg. 14, Nr. 24, S. 657 ff.; *Dinter, H.*, Zur Frage der Verwendung von Stichprobenverfahren im Rahmen von Prüfungshandlungen, Die Wirtschaftsprüfung, Februar 1962, Nr. 4, S. 85 ff.; *Rackles, R.*, a. a. O.

renden dieser Fachgebiete eine Einführung in die grundlegenden statistischen Stichprobenpläne und einige der statistischen Stichprobentechniken geben, die bei internen und externen Rechnungsprüfungen nützlich sein könnten¹⁾“.

In Europa, wo die Diskussion die in Amerika beobachtete Weite noch nicht erreicht hat, beschäftigen sich mehrere nationale Berufsorganisationen und die Union Européenne des Experts Comptables Economiques et Financiers (UEC) mit diesen Problemen. Aber diese Arbeiten sind bisher noch nicht weit fortgeschritten, nicht zuletzt, weil entsprechend vorgebildete Kräfte selten sind.

In den nachfolgenden Ausführungen soll nun zunächst ein kurzer Abriß über die Abschlußprüfung, insbesondere über die gegenwärtige Prüfungstechnik und anschließend ein Überblick über die Grundbegriffe der Stichprobentheorie gegeben werden. Im Hauptteil wird sich die Arbeit dann mit der Frage befassen, ob und gegebenenfalls auf welchen Prüfungsgebieten und in welchem Umfang die Stichprobentheorie zur Verbesserung der gegenwärtigen Prüfungsmethoden beizutragen vermag.

II. Abriß über die Jahresabschlußprüfung, insbesondere über die Prüfungstechnik

1. Grundlagen der Jahresabschlußprüfung

Die Untersuchung wird sich mit der *Jahresabschlußprüfung nach aktienrechtlichen Grundsätzen* befassen, da sie hinsichtlich Prüfungsziel und Prüfungstechnik als der Prototyp aller gesetzlichen und freiwilligen Jahresabschlußprüfungen angesehen werden kann²⁾.

Die Jahresabschlußprüfung ist die umfassende und grundlegende Gesamtprüfung der Rechnungslegung eines Wirtschaftsunternehmens für eine bestimmte Periode, nämlich das Geschäftsjahr³⁾. Sie erstreckt sich auf die Einhaltung der Grundsätze der materiellen und formellen Ordnungsmäßigkeit der Rechnungslegung, insbesondere auf die Einhaltung der hierzu bestehenden gesetzlichen Vorschriften⁴⁾. Sie besteht in der gewissenhaften, unparteiischen und sachverständigen Beurteilung der das Vermögen oder den Erfolg beeinflussenden wirtschaftlichen Vorgänge und ihrer rechnungsmäßigen Behandlung sowie in der Auswertung der Prüfungsergebnisse⁵⁾.

1) a. a. O., Vorwort, S. VII

2) Hintner, a. a. O., S. 17 und 61 f.; *Auffermann, J. D.*, Jahresabschlußprüfung, Handwörterbuch der Betriebswirtschaft (HdBW), 3. Aufl., Stuttgart 1958, 8. Bd., Spalte 2897; *Baudisch, R.*, Die Funktion des Abschlußprüfers, Wirtschaftsprüfung als Element gesunder Wirtschaftsentwicklung, Düsseldorf 1955, S. 24 f.; WP-Handbuch 1959, Hrsg. vom Institut der Wirtschaftsprüfer, Düsseldorf 1959, S. 105.

3) *Adler/Düring/Schmaltz*, a. a. O., § 135, Tz. 18; *Auffermann*, a. a. O., HdBW, Spalte 2897.

4) *Adler/Düring/Schmaltz*, a. a. O., § 135, Tz. 18, 93, 114; WP-Handbuch 1959, a. a. O., S. 105; *Auffermann*, a. a. O., HdBW, Spalte 2897.

5) *Adler/Düring/Schmaltz*, a. a. O., § 141, Tz. 3; *Berg, K.*, Wie wird geprüft?, 6. Aufl., Bonn 1949, S. 7, 15 f.

Ihre gesetzliche Grundlage findet die Jahresabschlußprüfung in den §§ 135 bis 142 Aktiengesetz (AktG).

Ihr Zweck ist die Sicherung der Einhaltung der *Rechnungslegungsvorschriften* für die Aktiengesellschaften. Sie sind in den §§ 127 bis 134 AktG niedergelegt. Die Vorschriften über die Gliederung der Gewinn- und Verlustrechnung sind durch das „Gesetz über die Kapitalerhöhung aus Gesellschaftsmitteln und über die Gewinn- und Verlustrechnung“ vom 23. 12. 1959 wesentlich erweitert worden. Ergänzend zu den aktienrechtlichen Vorschriften sind die Vorschriften der §§ 38 bis 42 Handelsgesetzbuch (HGB) und die Grundsätze ordnungsmäßiger Buchführung und Bilanzierung zu beachten.

Diese Vorschriften haben teilweise formellen, teilweise materiellen Inhalt. Von *formeller Bedeutung* sind z. B. die Vorschriften über den Gebrauch einer lebenden Sprache, über das Verbot von Rasuren, das Gebot, freie Stellen in den Büchern zu entwerfen und das Gebot, die Unterlagen geordnet aufzubewahren, ferner verschiedene Grundsätze ordnungsmäßiger Buchführung, wie das Belegprinzip, das Gebot, die Geschäftsvorfälle laufend zeitlich und sachlich geordnet zu verbuchen usw.

Diese Formvorschriften dienen der Verständlichkeit und Sicherheit der Rechnungslegung.

Zu den *materiellen Vorschriften* gehören vor allem diejenigen über die Bewertung der Aktiva und Passiva (§ 133 AktG) und die Gliederung der Bilanz (§ 131 AktG) und der Gewinn- und Verlustrechnung (§ 132 AktG).

Die *Bewertungsvorschriften* des Handelsrechts, besonders des Aktienrechts, bezwecken die vorsichtige Ermittlung des Jahreserfolges, insbesondere durch Verhinderung des Ausweises noch nicht realisierter Gewinne, indem sie Höchstwerte für die Bewertung des Anlage- und des Umlaufvermögens festsetzen.

Die *Gliederungsvorschriften* verlangen eine übersichtliche und klare Aufgliederung der Aktiva und Passiva in der Bilanz sowie eine weitgehende Aufgliederung der Aufwands- und Ertragsquellen in der Gewinn- und Verlustrechnung. Sie verlangen neuerdings den Ausweis der Bruttobeträge und verbieten die bis zum Erlaß des Gesetzes über die Kapitalerhöhung aus Gesellschaftsmitteln und über die Gewinn- und Verlustrechnung vom 23. 12. 1959 erlaubten Saldierungen.

Die materiellen Vorschriften betreffen also den Inhalt des Jahresabschlusses.

Zweck dieser Vorschriften ist es in erster Linie, den *Aktionären und Gläubigern* Einblick in die wirtschaftlichen Verhältnisse des Unternehmens, in seine Ertrags-, Vermögens-, Finanz- und Liquiditätslage zu geben, damit sie ihre Interessen, die nicht immer mit den Interessen des Unternehmens konform gehen müssen, wahrnehmen und nötigenfalls schützen können. Darüber hinaus sollen die Organe der Unternehmung und in eingeschränktem Maße die Öffentlichkeit zuverlässig über die Lage des Unternehmens unterrichtet werden.

In diesem Zusammenhang muß hervorgehoben werden, daß die Bewertungsvorschriften für das Anlage- sowie für das Umlaufvermögen (§ 133 Abs. 1 AktG) Höchstwerte vorschreiben, die nicht überschritten, wohl aber unterschritten werden dürfen, da die Bildung *stiller Reserven* nach den gegenwärtigen gesetzlichen

Vorschriften in gewissem Rahmen erlaubt ist¹⁾. Ähnlich können auch bei der Bewertung von Rückstellungen und Wertberichtigungen stille Reserven zulässigerweise gebildet werden. Das bedeutet, daß in vielen Fällen nicht von *einem* allein richtigen Bilanzansatz gesprochen werden kann, sondern daß innerhalb eines gewissen *Spielraums* alle Werte als gesetzlich zulässig anzusehen sind.

2. Prüfungstechnik

Prüfung ist immer die Wertung des Vorgangs, dem die Prüfung gilt, hinsichtlich seines Inhalts und der Form, in welcher er dokumentiert wird²⁾. Wichtigste Voraussetzungen einer zutreffenden Wertung sind genügende Sachkenntnis im allgemeinen und Würdigung des Vorgangs im Rahmen des betreffenden Unternehmens³⁾.

Zur Bewältigung des Prüfungsstoffs bedient sich der Prüfer einer bestimmten *Prüfungstechnik*. Sie besteht aus den Prüfungshandlungen und den Prüfungs wegen⁴⁾.

Alle *Prüfungshandlungen* sind grundsätzlich im Hinblick auf die Feststellung der Gesetz- und Ordnungsmäßigkeit der Rechnungslegung vorzunehmen und die festgestellten Tatsachen müssen für das Endergebnis der Prüfung verwertbar sein, auch wenn sie zunächst nur Teilgebiete oder Einzelheiten der gesamten Rechnungslegung zum Gegenstand haben⁵⁾.

Die Prüfungshandlungen lassen sich im wesentlichen nach drei Handlungsarten, nämlich Vergleichen, Rechnen und Aufnehmen, ordnen.

Ein *Vergleich* kann zwischen Aufzeichnungen von Vorgängen und Sachverhalten und den ihnen zugrunde liegenden tatsächlichen Vorgängen bzw. Sachverhalten und zwischen mehreren Aufzeichnungen angestellt werden.

Die im Rahmen der Abschlußprüfung anzuwendenden *Rechenoperationen* lassen sich gliedern in solche mit reinem Wiederholungscharakter und solche mit eigener selbständiger Prägung. Eine weitere Unterscheidung der ersten ist die nach genauen und nach überschlägigen Wiederholungen. Rechenoperationen selbständiger Prägung liefern das Prüfungsergebnis aufgrund anderer Berechnungsweisen als der zu prüfenden und haben entweder reinen Kontrollcharakter („Kontrollzahlen“) oder außer dem Kontrollcharakter auch eine Bedeutung im Rahmen des zu erstattenden Prüfungsberichts (z. B. Entwicklungsreihen).

Eine weitere wichtige Art von Prüfungshandlungen ist die *Aufnahme* (Inventur). Das geschieht bei körperlichen Beständen durch Messen, Zählen, Wiegen, bei Rechten und immateriellen Werten, insbesondere Forderungen und Verpflichtungen, durch ihre listenmäßige Erfassung.

1) RGZ Bd. 116, S. 119; ferner RG vom 15. 10. 1937 (II 61/37); siehe dazu ferner WP. Jahrbuch 1959, a. a. O., S. 488 f. und Adler/Düring/Schmaltz, a. a. O., § 129, Tz. 71 ff. und § 133, Tz. 9 ff.

2) Hintner, a. a. O., S. 63.

3) Adler/Düring/Schmaltz, a. a. O., § 135, Tz. 116 (vgl. oben, S. 1).

4) Hintner, a. a. O., S. 62 f.

5) Adler/Düring/Schmaltz, a. a. O., § 135, Tz. 122 und 144; Hintner, a. a. O., S. 69

Die bei der Abschlußprüfung vorkommenden *Prüfungswege*¹⁾ können gegliedert werden nach

- a) dem Prüfungsinhalt (formelle und materielle Prüfungen);
- b) der Prüfungsrichtung (progressive und retrograde Prüfungen);
- c) dem Prüfungsausmaß (lückenlose und stichprobenweise Prüfungen);
- d) dem Prüfungsobjekt (Prüfung einzelner Geschäftsvorfälle — Prüfung der Organisation des Rechnungswesens).

Das geläufigste Einteilungsprinzip ist das in *formelle und materielle Prüfungen*. Es hat sich historisch entwickelt. Die Geschichte der Buchprüfung und der Berufsprüfung der Wirtschaftsprüfer wird in Deutschland beherrscht von der Verlagerung des Schwergewichts von der rein formellen zur überwiegend materiellen Prüfung²⁾.

Trotz zahlreicher Abgrenzungsversuche wird der Wert dieser Gliederung nicht widerspruchslos anerkannt³⁾. Überwiegend versteht man unter *formeller* Prüfung diejenigen Prüfungshandlungen, welche sich auf den formal-rechnerischen Inhalt der Buchführung beziehen („äußere Ordnungsmäßigkeit“⁴⁾). Die *materielle* Prüfung hingegen erfaßt, prüft und würdigt kritisch die zugrunde liegenden wirtschaftlichen Vorgänge und Zusammenhänge hinsichtlich ihrer inhaltlich richtigen Darstellung und deren wirtschaftlicher Berechtigung⁵⁾.

Hiernach lassen sich schon rein begrifflich formelle und materielle Prüfungen *nicht eindeutig gegeneinander abgrenzen*. Das ist in der Prüfungspraxis noch viel schwieriger, weil formelle und materielle Prüfungshandlungen eng miteinander verbunden sind und häufig ineinander übergehen⁶⁾. Nach der vorstehenden Definition soll z. B.⁷⁾ bei der Prüfung eines Inventars die Feststellung der Einhaltung der Inventurrichtlinien eine formelle Prüfung, das Nachzählen, Wiegen, Abmessen usw. der einzelnen Bestände und der Vergleich des Ergebnisses mit den Aufzeichnungen eine materielle Prüfung sein; die Prüfungen der Additionen und Unterschriften auf den Urschriften und der Übertragungen von diesen auf die Reinschrift sollen formellen Charakter haben, die Prüfung der angesetzten Einzelpreise wiederum eine materielle und die Prüfung der Ausrechnung des Postenwertes (Einzelpreis mal Menge) schließlich eine formelle Handlung sein. Auch die Abstimmung der Inventurergebnisse mit den Lagerkarteien (Differenzen der Mengeninventur) einerseits und die Abstimmung der Inventurergebnisse mit den Zahlen der Finanzbuchhaltung (Differenzen der Wertinventur) andererseits rechnen nach den obigen Definitionen zu den formellen Prüfungsakten, während die Analyse dieser Differenzen materielle Prüfungshandlungen darstellen.

Eine solche Art der Unterscheidung zwischen formeller und materieller Prüfung kann sachlich nicht befriedigen. Die Unterscheidung von formeller und materieller

1) Vgl. *Hintner*, a. a. O., S. 63 ff.

2) *Minz*, a. a. O., S. 90.

3) *Minz*, a. a. O., S. 89, dazu führt *Minz* eine Reihe von Belegstellen in Fußnote 1, S. 89, an.

4) *Minz*, a. a. O., S. 89.

5) *Hintner*, a. a. O., S. 64; *Minz*, a. a. O., S. 89.

6) *Hintner*, a. a. O., S. 64.

7) Siehe *Minz*, a. a. O., S. 89 f.

Prüfung wird jedoch *eindeutig und sinnvoll*, wenn man die oben getroffene Unterscheidung der Rechnungslegungsvorschriften in solche formellen und solche materiellen Charakters auf die Prüfung überträgt¹⁾.

Dann sind *formelle Prüfungen* alle diejenigen, die die Einhaltung der Vorschriften zur *Sicherheit und Verständlichkeit der Rechnungslegung* betreffen. Hierzu rechnen also die oben bereits angeführten *handelsrechtlichen Bestimmungen*, wie die Gebote des Gebrauchs einer lebenden Sprache und nicht löscharer Schreibmittel, die Verbote von Rasuren und leeren Zwischenräumen, die Pflicht zur Numerierung und geordneten Aufbewahrung der Belege, zur fortlaufenden Eintragung der Geschäftsvorfälle in die Bücher usw.²⁾. Hierzu gehören dann aber auch die *unternehmensinternen Ordnungsvorschriften*, soweit sie den Zweck haben, die Verständlichkeit und Sicherheit der Rechnungslegung zu gewährleisten. Das sind in erster Linie die Vorschriften des zwar allgemein angewandten, aber gesetzlich nicht vorgeschriebenen Systems der doppelten Buchführung (Doppik), also die doppelte Verbuchung jedes Geschäftsvorfalles im Soll und im Haben, ferner die Einhaltung der Kontrollvorschriften im Waren-, Leistungs- und Zahlungsverkehr usw.

Die *materiellen Prüfungen* sind dann alle diejenigen, die sich auf die Richtigkeit des *Inhalts des Jahresabschlusses*, d. h. sowohl der Bilanz als auch der Gewinn- und Verlustrechnung beziehen, also in erster Linie die Prüfung der Bewertungs- und Gliederungsvorschriften. Auf sie wurde bereits oben eingegangen. Dann sind aber z. B. Rechenfehler in den Belegen oder Übertragungsfehler vom Beleg ins Konto keine Formalfehler, sondern materielle Fehler, weil sie den Inhalt des Jahresabschlusses unzulässig verändern. Ein Rechen- oder Übertragungsfehler in den Konten ist dagegen ein Verstoß gegen die Regeln der Doppik, der die Summengleichheit stört. Er verhindert bereits die Aufstellung eines formell ordnungsmäßigen Jahresabschlusses und muß bei der Prüfung der Doppik auftauchen und eliminiert werden.

Im Laufe dieser Untersuchung wird sich zeigen, daß diese Einteilung sich unter statistischen Gesichtspunkten geradezu aufdrängt und zu brauchbaren Ergebnissen führt.

Ein zweites Einteilungsprinzip der Prüfungswege ist das nach der *Prüfungsrichtung*. Bei der *progressiven* Prüfungsweise führt der Prüfungsweg vom Beleg als Buchungsgrundlage über die Grundbücher, Journale, das Hauptbuch hinweg zur Bilanz bzw. Gewinn- und Verlustrechnung. Die *retrograde* Prüfung geht den umgekehrten Weg von der Bilanz bzw. Gewinn- und Verlustrechnung zum Beleg. Selten wird jedoch vom Prüfer einer dieser Wege ganz oder im Laufe der gesamten Prüfung allein beschritten, sondern meist wird mitten in eine Buchungskette hineingegriffen und dann nur nach der einen oder nach beiden Richtungen geprüft.

Eine spezielle Form der retrograden Prüfung ist die sogenannte „*Wurzelstichprobe*“. Damit wird von der Prüfungspraxis eine Stichprobe bezeichnet, bei der nicht beim Beleg Halt gemacht, sondern der Posten in allen Einzelheiten und unter

¹⁾ Siehe dazu *Marcantonio, A.*, Die Durchführung der Revision, Probleme neuzeitlicher Unternehmensführung, Arbeitsunterlagen für den IV. Kongreß der UEC vom 18.—21. 9. 1961 in Zürich, S. 137 ff.

²⁾ Im einzelnen s. die Bestimmungen des § 162 Abgabenordnung, die als Grundsätze ordnungsmäßiger Buchführung auch handelsrechtliche Bedeutung haben.

Heranziehung aller zugehörigen aktenmäßigen Unterlagen und mit dem Prüfungsgegenstand in Zusammenhang stehenden Nachweise geprüft wird, wobei auch die jeweiligen Gegenbuchungen in die Prüfung einzubeziehen sind¹⁾. Nur eine solche Prüfung erfüllt den „Grundsatz, daß jede einzelne Prüfungshandlung als solche in sich lückenlos sein muß. Posten ohne Einsichtnahme aller Belege und Unterlagen zu „prüfen“, ist falsch und zwecklos²⁾“.

Die Einteilung in *lückenlose* und *stichprobenweise* Prüfungen ist im Rahmen dieser Arbeit von besonderem Interesse. Eine *lückenlose Prüfung* liegt vor, wenn innerhalb eines zeitlich, räumlich und sachlich bestimmten Bereichs die gesamten in Frage kommenden Vorgänge oder Tatbestände geprüft werden. Bei der *stichprobenweisen Prüfung* werden grundsätzlich nicht alle in den Bereich fallenden Vorgänge oder Tatbestände geprüft, sondern es wird eine Auswahl getroffen, d. h. eine Teilmasse herausgegriffen.

Es bedarf kaum einer besonderen Erklärung, daß mit Zunahme der Unternehmensgröße der Umfang der Geschäftstätigkeit und damit des Prüfungsstoffes sowohl nach Breite (Außenumsätze) wie nach Tiefe (Ausgestaltung des innerbetrieblichen Rechnungswesens) ein solches Ausmaß angenommen hat, daß eine lückenlose Prüfung praktisch in angemessener Zeit und zu angemessenen Kosten nicht bewältigt werden kann.

Die lückenlose Prüfung bleibt daher in der Regel auf Unternehmen beschränkt, deren Geschäftsvorfälle verhältnismäßig gering an Zahl sind. Dabei kann es sich jedoch — z. B. bei Holding-Gesellschaften — um sehr bedeutende Unternehmen handeln, die in vergleichsweise wenigen Transaktionen unter Umständen wertmäßig bedeutende Umsätze tätigen³⁾.

Je umfangreicher das in die Stichprobe einbezogene Material ist, desto häufiger wird es vorkommen, daß die letzten Schritte der Verfolgung der ausgewählten Posten bis in die Bilanz und Gewinn- und Verlustrechnung über Buchungswege führen, die an Hand eines bereits geprüften Postens schon einmal einer Prüfung unterzogen wurden. Es wäre unsinnig, dieselben Gegenstände nochmals einer Prüfung zu unterziehen, da ihr Ergebnis ja bereits feststeht.

In den Vordergrund des Prüferinteresses tritt in solchen Fällen die Prüfung der vorhandenen *Prüfungs- bzw. Kontrolleinrichtungen* des geprüften Unternehmens, insbesondere aller jener Einrichtungen und Maßnahmen, die die Prüfung bzw. Kontrolle der Organisation der Rechnungslegung des Unternehmens zum Ziel haben⁴⁾. Diese Einrichtungen und Maßnahmen können so weit entwickelt werden, daß sie ein mehr oder minder geschlossenes System (innerbetriebliches Kontrollsystem) darstellen, das „den Organisationsplan sowie alle aufeinander abgestimmten Methoden und Maßnahmen umfaßt, die in einem Unternehmen angewandt werden⁵⁾“ mit dem Zweck, die genaue Erfassung aller Geschäftsvorfälle zu

1) *Hintner*, a. a. O., S. 66 f.; ebenso *Adler/Düring/Schmaltz*, a. a. O., § 135, Tz. 122.

2) *Hintner*, a. a. O., S. 67.

3) *Hintner*, a. a. O., S. 65.

4) *Pougin, E.*, Die Berücksichtigung des internen Kontrollsystems als Grundlage ordnungsmäßiger Abschlußprüfung, Düsseldorf 1959, S. 77, ähnlich *Hintner*, a. a. O., S. 68.

5) American Institute of Certified Public Accountants, Internal Control, Special Report by the Committee on Auditing Procedure, 1949, S. 6.

bewirken und damit die Richtigkeit der Rechnungslegung zu gewährleisten¹⁾. Ferner soll durch diese Maßnahmen das Vermögen gegen Unterschlagungen und andere Unregelmäßigkeiten gesichert, die Einhaltung der Geschäftspolitik gefördert und der betriebliche Wirkungsgrad erhöht werden.

Das beste, in sich geschlossene System ist aber wirkungslos, wenn die Einrichtungen nicht gehandhabt und die Maßnahmen nicht durchgeführt werden. Der Abschlußprüfer wird sich also ein Urteil darüber bilden müssen, inwieweit er sich bei seiner Prüfung auf die *Wirksamkeit des innerbetrieblichen Kontrollsystems* verlassen und es sich zunutze machen kann²⁾.

Das Gesamturteil des Prüfers über die Geschlossenheit des innerbetrieblichen Kontrollsystems einschließlich des Erkennens seiner Mängel und über die tatsächliche Wirksamkeit der Kontrollen sollen nach moderner Anschauung die *Grundlage für die Entscheidung des Prüfers* über Auswahl und Umfang seiner eigentlichen Abschlußprüfungen liefern. So schreibt *Meier*³⁾:

„Sowohl die Stichprobenprüfung als auch die Nachprüfung der Arbeiten der Hausrevision sind schon nicht mehr unmittelbar, sondern nur noch mittelbar auf das eigentliche Prüfungsziel gerichtet. Sie dienen nämlich in erster Linie dazu, ein Urteil über die Zuverlässigkeit der Organisation des Rechnungswesens und seiner Kontrolleinrichtungen zu gewinnen. Dabei kommt es weniger darauf an, in welchem Umfang das Rechnungswesen mechanisiert ist, als vielmehr darauf, in welchem Maße es sich zwangsläufig abwickelt und kontrolliert. Wenn und insoweit nach der Übernahme des Beleginhalts auf eine Maschine, z. B. eine elektronische Recheneinheit, die aufbereitenden und auswertenden Arbeiten zwangsläufig durch die Maschine erledigt werden, braucht sich m. E. die Prüfung des Abschlußprüfers nur noch auf den Beleginhalt, die Belegübernahme und darauf zu erstrecken, daß die der Maschine gegebene Arbeitsanweisung (Programmierung) richtig ist sowie daß die Maschine fehlerfrei arbeitet. So wird die Abschlußprüfung, soweit sie sich auf die Buchführung bezieht, mit der fortschreitenden Automation des Rechnungswesens in wachsendem Maße zu einer Organisationsprüfung, die allerdings nicht auf die Zweckmäßigkeit dieser Organisation, sondern auf die Zuverlässigkeit ihrer Ergebnisse gerichtet ist.“

Daraus folgt, daß dort, wo das innerbetriebliche Kontrollsystem gut ist und gut funktioniert, wenig Stichproben und dort, wo es schlecht ist oder schlecht funktioniert, mehr Stichproben nötig sind.

Pougin stellt zusammenfassend folgendes fest⁴⁾:

„Der Stand des internen Kontrollsystems ist zwar kein exakter Maßstab — den gibt es nur im Bereich der exakten Wissenschaften — für die Bemessung des Prüfungsumfangs. Er ist aber eine Bemessungsgrundlage, auf der sich über das pflichtgemäße Ermessen des einzelnen Wirtschaftsprüfers hinaus eine einheitliche Berufsauffassung über den Prüfungsumfang wird bilden lassen, und zwar am ehesten deshalb, weil die grundlegenden Merkmale des internen Kontrollsystems

¹⁾ *Pougin*, a. a. O., S. 50.

²⁾ *Pougin*, a. a. O., S. 81.

³⁾ *Meier*, A., Rechtliche und tatsächliche Gestaltung der Jahresabschlußprüfung von Aktiengesellschaften, ZfhF, N. F., Jg. 9, 1957, S. 399.

⁴⁾ a. a. O., S. 106.

in allen Unternehmen die gleichen sind. Außerdem läßt sich diese Bemessungsgrundlage, wie die Ansätze in den Vereinigten Staaten zeigen, durch die *Anwendung statistischer Methoden*¹⁾ in zunehmendem Maße verfeinern. Damit würde der Prüfungsumfang der mathematischen Behandlung zugänglich.“

Bevor die Möglichkeit der Anwendung statistischer Methoden bei der Abschlußprüfung untersucht wird, sollen zunächst kurz einige Grundbegriffe der Stichprobentheorie erläutert werden.

III. Grundbegriffe der Stichprobentheorie²⁾

Was man in der Prüfungspraxis als Stichprobe bezeichnet, ist nicht identisch mit dem, was man in der Statistik unter einer Stichprobe versteht. Beide Begriffe lassen sich aber unter dem Oberbegriff *Teilerhebungen* zusammenfassen. In beiden Fällen sollen über eine zeitlich, räumlich und sachlich begrenzte Menge, die *Grundgesamtheit*, aufgrund einer herausgegriffenen Teilmenge, der *Stichprobe im weiteren Sinn*, Aussagen gemacht werden.

Die der Grundgesamtheit angehörenden Elemente sind die *Untersuchungseinheiten*. Die interessierenden Eigenschaften der Untersuchungseinheiten nennt man *Untersuchungsmerkmale*. Grundgesamtheit und Untersuchungsmerkmale werden durch das Untersuchungsziel bestimmt und müssen ihm soweit als irgend möglich entsprechen.

Man unterscheidet kategoriale oder qualitative und meßbare oder quantitative Untersuchungsmerkmale. *Qualitative Merkmale* sind z. B. die Fehlerfreiheit oder Fehlerhaftigkeit der Untersuchungseinheit, die Zugehörigkeit einer Einheit zu einer bestimmten Wertklasse, z. B. Klasse I: unter DM 1000,—, Klasse II: DM 1000,— bis unter DM 10 000,—, Klasse III: DM 10 000,— und darüber. Die Untersuchung von Einheiten auf ein qualitatives Untersuchungsmerkmal hat die Ermittlung der *Zahl* der Einheiten mit diesem Merkmal in der Grundgesamtheit oder des *Anteils* dieser Einheiten am Umfang der Grundgesamtheit zum Ziel. *Quantitative Merkmale* sind z. B. das Gewicht, der Rauminhalt, der Wert. Hier interessieren die *Gesamtmenge* (kg, cbm usw.) bzw. der *Gesamtwert* der Grundgesamtheit oder die *durchschnittliche Menge* oder der *durchschnittliche Wert* je Untersuchungseinheit. Quantitative Merkmale lassen sich durch Klassenbildung in qualitative Merkmale überführen, z. B. der Wert in DM in Klassen wie oben.

Der Unterschied zwischen der Stichprobe im Sinne der Prüfungspraxis und der statistischen Stichprobe besteht in der Art der Auswahl der Teilmenge.

Die *prüferische Praxis* bedient sich bisher fast ausschließlich der *bewußten Auswahl* dieser Teilmenge, d. h. einer Methode, bei der die einzelnen in die Untersuchung einzubeziehenden Einheiten ausschließlich nach dem subjektiven prüferischen Ermessen ausgewählt werden. *Stichproben im statistischen Sinn* sind demgegenüber Teilerhebungen, bei denen der *Zufall* darüber entscheidet, welche

1) Hervorhebung vom Zitierenden

2) In Anlehnung an *Kellerer, H.*, Theorie und Technik des Stichprobenverfahrens, Einzelschriften der Deutschen Statistischen Gesellschaft, Nr. 5, 2. Aufl., München 1953, und ders., Statistik im modernen Wirtschafts- und Sozialleben, Reinbek bei Hamburg 1960.

Einheiten in die Auswahl kommen. Dabei gilt als Zufall nur das, was bei einem entsprechenden Zufallsmodell der Wahrscheinlichkeitslehre vorkommen kann¹⁾. Ein Beispiel für ein solches Zufallsmodell ist das Ziehen von Losen aus einer Urne (Lotteriezählung). Die ausgewählten Einheiten bezeichnet man als *Stichprobeneinheiten*. In den folgenden Ausführungen werden unter Stichproben nur noch Stichproben im statistischen Sinn verstanden.

Das Wesen der Zufallsauswahl ist, daß jede Untersuchungseinheit eine bestimmte, berechenbare und von Null verschiedene Wahrscheinlichkeit hat, in die Stichprobe gezogen zu werden. Dadurch werden die Wahrscheinlichkeiten bestimmter *Stichprobenergebnisse* und ihrer *Zufallsfehler*, d. s. die Differenzen zwischen den Stichprobenergebnissen und dem Ergebnis der mit gleicher Sorgfalt durchgeführten *Vollerhebung*²⁾, berechenbar. Mithin können Stichprobenverfahren grundsätzlich so geplant werden, daß die Zufallsfehler mit bestimmter Wahrscheinlichkeit eine vorgegebene Größe nicht überschreiten. Unter verschiedenen in Frage kommenden Stichprobenverfahren kann ferner dasjenige bestimmt werden, das am rationellsten ist, d. h. bei dem der erreichte Genauigkeitsgrad im günstigsten Verhältnis zu dem für die Erhebung erforderlichen Aufwand steht.

Zufallsauswahl und bewußte Auswahl haben gemeinsam, daß sie hinsichtlich der Untersuchungsmerkmale möglichst repräsentativ für die Grundgesamtheit sein sollen, die Stichprobenverfahren haben aber gegenüber anderen Arten von Teilerhebungen, wie der bewußten Auswahl und der Auswahl aufs Geratewohl, den Vorzug, mit rationellen Methoden quantitative Aussagen zu erlauben, während bei den genannten anderen Teilerhebungen nur subjektive Urteile möglich sind.

Jede *Stichprobenerhebung* erfolgt nach einem *Stichprobenplan*. Dieser regelt das Auswahlverfahren und das Verfahren zur Ermittlung der gesuchten Ergebnisse. Die *Auswahl* der Stichprobeneinheiten erfolgt an Hand der *Auswahlgrundlagen*. Sie sollen ein schnelles Auffinden und die eindeutige Identifizierung der ausgewählten Untersuchungseinheiten ermöglichen. Art und Umfang der Auswahlgrundlagen und die Anordnung der Einheiten in ihnen sind wesentliche Faktoren für die Planung des Auswahlverfahrens. Als Auswahlgrundlagen kommen u. a. Lagepläne, Kontenverzeichnisse, Debitorenlisten, Inventurlisten etc. in Frage.

Bei der *Auswahl* der Stichproben können verschiedene *Verfahren* angewandt werden. Die wichtigsten Typen von Auswahlverfahren sind das einfache, das geschichtete, das Klumpen- und das mehrstufige Stichprobenverfahren. Zu jeder Verfahrensart gehört ein „Urnenmodell“. Die in der Praxis zu beobachtende Vielfalt der Auswahlverfahren beruht darauf, diese Typen teilweise miteinander kombinieren zu können.

Beim *einfachen Stichprobenverfahren* (uneingeschränkte Zufallsauswahl) enthält die Urne ebensoviel Lose, wie die Grundgesamtheit Untersuchungseinheiten. Jedes Los repräsentiert eine ganz bestimmte Einheit. Von den gut durchgemischten

¹⁾ *Anderson, O.*, Probleme der statistischen Methodenlehre in den Sozialwissenschaften, 3. Aufl., Würzburg 1957, S. 92.

²⁾ *Deming, W. E.*, Sample Design in Business Research, New York—London 1960, S. 50.

Losen werden so viele der Urne entnommen, wie die Stichprobe Einheiten umfassen soll. Die den ausgewählten Losen entsprechenden Untersuchungseinheiten sind die Stichprobeneinheiten. Bei der einfachen Stichprobe hat jedes in der Urne befindliche Los die gleiche Chance, in die Stichprobe zu kommen.

Für die praktische Anwendung ist dieses Verfahren allerdings zu umständlich. Es gibt aber eine Reihe anderer Auswahlmethoden, die diesem Urnenmodell praktisch gleichwertig sind. Hierzu zählen z. B. die Auswahl an Hand der „Zufallszahlentabellen“, das systematische Auswahlverfahren, das Schlußziffernverfahren, die Buchstabenauswahl.

Die „Zufallszahlentabellen“ enthalten z. B. 10 000 oder 15 000 Ziffern in völlig zufälliger, z. B. durch Lotterieziehung gewonnener, Reihenfolge und stellen gewissermaßen eine Urne auf Vorrat dar. Ist nun die Stichprobe aus einer Grundgesamtheit von z. B. 6000 durchnummerierten Einheiten zu ziehen, so beginnt man an irgendeiner Stelle der Tabelle von den aufeinanderfolgenden Ziffern je vier zu einer Zahl zusammenzufassen, bis soviel Zahlen gebildet worden sind, wie Stichprobeneinheiten ausgewählt werden sollen. Die so gebildeten Zahlen stellen die Nummern der in die Stichprobe einzubeziehenden Einheiten dar. Zahlen über 6000 bleiben unberücksichtigt.

Um ein *systematisches Auswahlverfahren* handelt es sich, wenn z. B. bei einem relativen Stichprobenumfang von $\frac{1}{50}$ des Umfangs der Grundgesamtheit unter den ersten 50 Einheiten eine, z. B. die neunte, (zufällig) herausgegriffen und dann jede 50. weitere (also die 59., 109., 159. usw.) in die Stichprobe gezogen wird. Dieses Verfahren ist jedoch nur anwendbar, wenn keine Übereinstimmung zwischen dem Stichprobenintervall und entsprechenden Perioden unter den Untersuchungseinheiten, die für die Untersuchungsmerkmale bedeutsam sind, besteht.

Das *Schlußziffernverfahren* setzt eine laufend durchnummerierte Grundgesamtheit voraus. Aus ihr werden, wenn ein relativer Stichprobenumfang von z. B. 10% auszuwählen ist, alle Einheiten in die Stichprobe gezogen, deren Schlußziffer beispielsweise 6 ist. Auch hier darf das Stichprobenintervall nicht mit Perioden unter den Untersuchungseinheiten übereinstimmen, die sich auf die Untersuchungsmerkmale auswirken.

Die *Buchstabenauswahl* läßt sich grundsätzlich bei Grundgesamtheiten anwenden, deren Einheiten nach Namen unterschieden werden können. Es sind dann alle Einheiten mit bestimmten Anfangsbuchstaben in die Stichprobe zu nehmen. Dieses Verfahren setzt voraus, daß kein Zusammenhang zwischen den Anfangsbuchstaben und den interessierenden Untersuchungsmerkmalen vorliegt.

Bei dem *geschichteten Stichprobenverfahren* wird eine heterogene Grundgesamtheit derart in Teilgesamtheiten (Schichten) untergliedert, daß diese in sich homogener sind, als die Grundgesamtheit insgesamt. In jeder Schicht erfolgt eine gesonderte Auswahl der Stichprobeneinheiten. Im zugehörigen Urnenmodell sind ebensoviel Urnen vorhanden wie Schichten. Aus jeder Urne wird eine Anzahl Lose gezogen. Hierbei ist — im Gegensatz zur einfachen Stichprobe — die Chance eines Loses, in die Stichprobe zu kommen, von Urne zu Urne verschieden. Jede Untersuchungseinheit hat indes eine berechenbare und von Null verschiedene Chance, in die Stichprobe zu gelangen.

Beim *Klumpenstichprobenverfahren* werden Klumpen von untereinander möglichst inhomogenen Untersuchungseinheiten gebildet. Anschließend wird eine bestimmte Anzahl der Klumpen ausgewählt und alle Einheiten dieser Klumpen in die Stichprobe gezogen. Das zugehörige Urnenmodell umfaßt soviel Urnen, wie Teilgesamtheiten (Klumpen) vorhanden sind, und jede Urne enthält die Lose der Einheiten einer bestimmten Teilgesamtheit. Von den Urnen wird eine Anzahl zufällig ausgewählt, und in die Stichprobe kommen alle diejenigen Einheiten, deren Lose sich in den ausgewählten Urnen befinden.

Das *mehrstufige Stichprobenverfahren* wird insbesondere bei organisatorisch unterteilten und räumlich weit gestreuten, umfangreichen Grundgesamtheiten angewandt. Der einfachste Fall eines solchen Verfahrens ist das zweistufige Verfahren. Dieses zeichnet sich dadurch aus, daß zunächst auf der ersten Auswahlstufe — wie beim Klumpenstichprobenverfahren — eine Anzahl von Teilgesamtheiten zufällig ausgewählt und sodann auf der zweiten Auswahlstufe aus jeder der ausgewählten Teilgesamtheiten wiederum eine Stichprobe gezogen wird. In dem zugehörigen Urnenmodell wird zunächst eine Auswahl unter den Urnen getroffen und sodann aus jeder ausgewählten Urne eine Anzahl von Losen gezogen. Dieses Modell läßt sich zu einem drei- oder vierstufigen Modell usw. erweitern. Beim mehrstufigen Verfahren wird also jeweils eine Anzahl von Untersuchungseinheiten zu einer *Auswahleinheit* zusammengefaßt. Alle Auswahleinheiten der gleichen Auswahlstufe bilden die *Auswahlgesamtheit* dieser Stufe.

Die Auswahl der Stichprobeneinheiten kann auch in mehreren aufeinanderfolgenden *Phasen* oder Stück für Stück, d. h. in Form der *Sequenzstichprobe*, vorgenommen werden. Die Auswahlverfahren können von Phase zu Phase verschieden sein, z. B. kann in der ersten Phase ein einfaches, in der zweiten Phase ein mehrstufiges Verfahren durchgeführt werden. Entsprechend unterscheiden sich dann die diesen zugehörigen Urnenmodelle.

Die Anwendung von geschichteten, von Klumpen- oder von mehrstufigen Stichprobenverfahren ermöglicht bei gleichen Kosten je nach der Struktur der Grundgesamtheit einen wesentlich kleineren Stichprobenumfang oder eine größere Genauigkeit des Ergebnisses als die Anwendung einer einfachen Stichprobe. Den relativen Umfang der Stichprobe zum Umfang der Grundgesamtheit bezeichnet man als *Auswahlsatz*.

Die Wahrscheinlichkeiten, in die Stichprobe zu kommen (Auswahlwahrscheinlichkeiten), die für die einzelnen Untersuchungseinheiten je nach Art des Auswahlverfahrens bestehen, bilden die Grundlagen für den *Formelapparat*, den die Stichprobentheorie entwickelt hat, um die jeweils interessierenden Beziehungen zwischen den Stichprobenergebnissen und den Strukturen der Grundgesamtheiten zu beschreiben.

Die Beziehungen können von zweierlei Art sein, je nach dem, welche *statistische Fragestellung* der Erhebung zugrunde liegt. Die Stichprobe kann

- a) der statistischen Prüfung von Hypothesen, die die Struktur der Grundgesamtheit betreffen;
- b) der Schätzung unbekannter Parameter der Grundgesamtheit dienen.

Als *Parameter* bezeichnet man charakteristische Werte (z. B. Anteilwerte, Mittelwerte, Gesamtwerte) der Grundgesamtheit.

Bei der *Prüfung einer Hypothese* über die Struktur einer Grundgesamtheit soll aufgrund einer Stichprobe entschieden werden, ob diese Hypothese angenommen werden kann oder abgelehnt (zurückgewiesen) werden muß.

Hierbei wird festgesetzt, daß eine Stichprobe bestimmten Umfangs höchstens eine vorgegebene Zahl fehlerhafter Einheiten enthalten darf, damit die Grundgesamtheit angenommen werden kann. Diese höchstzulässige Fehlerzahl heißt *Annahmezahl*. Übersteigt die vorgefundene Fehlerzahl die Annahmezahl, so wird die Grundgesamtheit abgelehnt. Umfang der Grundgesamtheit, Stichprobenumfang und Annahmezahl bilden zusammen den *Prüfplan*.

Die Stichprobe kann aber auch in mehreren Phasen gezogen werden. Bei mehrphasigen Annahmestichproben ist eine endgültige Entscheidung über Annahme oder Rückweisung der Grundgesamtheit unter Umständen bereits aufgrund eines Teils der Stichprobe möglich, wenn er auf einen besonders geringen oder besonders hohen Fehleranteil in der Grundgesamtheit schließen läßt. Sieht der Stichprobenplan mehrere Auswahlphasen vor, so muß der Prüfplan für jede Phase neben dem jeweiligen Stichprobenumfang nicht nur die Annahmezahl angeben, sondern auch die *Rückweisungszahl*, d. i. die kleinste Zahl fehlerhafter Stichprobeneinheiten, die zur Ablehnung der Grundgesamtheit führt. Ist die festgestellte Fehlerzahl in der ersten Phase größer als die Annahmezahl, aber kleiner als die Rückweisungszahl, so ist eine zweite Phase erforderlich. Die Auswahl ist so lange fortzusetzen, bis eine endgültige Entscheidung über Annahme oder Rückweisung der Grundgesamtheit fällt.

Ist der Prüfplan gegeben, so läßt sich für jeden beliebigen Fehleranteil in der Grundgesamtheit einerseits die Wahrscheinlichkeit dafür berechnen, daß die Zahl der in der Stichprobe vorgefundenen Fehler die Annahmezahl nicht überschreitet, d. h. daß die Grundgesamtheit angenommen wird — diese Wahrscheinlichkeit wird *Annahmewahrscheinlichkeit* genannt —, andererseits die Wahrscheinlichkeit dafür ermitteln, daß die Rückweisungszahl erreicht oder überschritten, m. a. W., daß die Grundgesamtheit abgelehnt wird — diese Wahrscheinlichkeit heißt *Rückweisungswahrscheinlichkeit* —.

Gibt man einen Fehleranteil als gerade noch zulässig vor, dann kann man bei gegebenem Prüfplan für diesen Fehleranteil — er wird *Rückweisungsgrenze* genannt — das *Annahmerisiko* berechnen, d. i. die Wahrscheinlichkeit, daß die Zahl der fehlerhaften Stichprobeneinheiten die Annahmezahl nicht überschreitet.

Ist ein „durchschnittlicher“ Fehleranteil vorgegeben, bei dem man die Grundgesamtheit anzunehmen wünscht, so kann man bei dem gegebenen Prüfplan für diesen Fehleranteil das *Rückweisungsrisiko* berechnen, d. i. die Wahrscheinlichkeit, daß die Zahl der fehlerhaften Stichprobeneinheiten mindestens gleich der Rückweisungszahl ist.

Bei Vorgabe eines Rückweisungsrisikos für den durchschnittlichen Fehleranteil und eines Annahmerisikos für die Rückweisungsgrenze ist es grundsätzlich möglich, den *Stichprobenumfang* sowie die *Annahme-* und *Rückweisungszahlen* zu ermitteln.

Ein Spezialfall der statistischen Hypothesenprüfung ist die *Entdeckungsstichprobe*¹⁾. Dabei handelt es sich um ein Stichprobenverfahren, mit dem eine bestimmte Zahl von Untersuchungseinheiten, die Träger selten vorkommender Merkmale sind, entdeckt werden, d. h. in die Stichprobe gelangen sollen. Die Wahrscheinlichkeit für das Auftauchen einer bestimmten Zahl dieser Einheiten ist berechenbar, wenn die Zahl dieser Einheiten in der Grundgesamtheit gegeben ist, und wird *Entdeckungswahrscheinlichkeit* genannt. Umgekehrt ist es möglich, den *Stichprobenumfang* zu berechnen, der erforderlich ist, damit diese Wahrscheinlichkeit eine bestimmte vorgegebene Größe erreicht.

Im Fall der *stichprobenweisen Schätzung* charakteristischer Werte der Grundgesamtheit werden die entsprechenden charakteristischen Werte der Stichprobe als *Schätzwerte* für die gesuchten Werte der Grundgesamtheit betrachtet. Die Schätzwerte sollen den gesuchten Werten möglichst nahekommen. Bei genügender Größe der Stichprobe und des nicht ausgewählten Restes der Grundgesamtheit läßt sich die Wahrscheinlichkeit berechnen, daß die Abweichung eines Schätzwertes vom gesuchten Werte eine bestimmte Größe nicht überschreitet. Den so begrenzten Bereich für die Schätzwerte nennt man *Vertrauensbereich*; die Wahrscheinlichkeit bezeichnet man als *Sicherheitsgrad*. Umgekehrt kann grundsätzlich der *Stichprobenumfang* ermittelt werden, der es gestattet, vorgegebene Größen des Sicherheitsgrades und des Vertrauensbereiches einzuhalten.

Nachfolgend werden zunächst der statistische Charakter der jeweiligen Prüfungsfragen und anschließend die mathematischen Zusammenhänge der verschiedenen Formeln, insbesondere die Wirkung der einzelnen in den Formeln vorkommenden Faktoren auf den Stichprobenumfang, untersucht.

¹⁾ Hierzu s. u. a. *Arkin, H.*, Discovery Sampling in Auditing, Journal of Accountancy (JofA), Februar 1961, S. 51 ff.

HAUPTTEIL

I. Gliederung der Prüfungsfragen

Nunmehr ist die Frage zu prüfen, ob und wo die Voraussetzungen für die Anwendung von Zufallsstichproben bei der Jahresabschlußprüfung gegeben sind. Ihre Beantwortung muß gesondert für die verschiedenen Prüfungsgebiete und Prüfungsfragen innerhalb der Jahresabschlußprüfung erfolgen.

Der *Prüfungsstoff* umfaßt die Prüfungsgebiete Buchführung, Inventar, Jahresabschluß (Bilanz und Gewinn- und Verlustrechnung) und Geschäftsbericht¹⁾.

Die *Prüfungsfragen* lassen sich auf folgende Grundfragen zurückführen:

- a) Ist der vorgelegte Prüfungsstoff vollständig?
- b) Sind die formellen Ordnungsvorschriften für die Rechnungslegung eingehalten?
- c) Ist die materielle Richtigkeit der Rechnungslegung gegeben?

Für die Prüfung des *Geschäftsberichts* scheiden Zufallsstichproben grundsätzlich aus, weil die in ihm enthaltenen Angaben und Erläuterungen keine gleichartigen Einheiten im Sinne der statistischen Theorie sind.

Die Prüfung der *Vollständigkeit* der Rechnungslegung hat sich sowohl auf die Vollständigkeit der Erfassung der Bestände (Inventar) als auch der Geschäftsvorfälle (Buchführung und Jahresabschluß) zu erstrecken.

Die Prüfung der *formellen Ordnungsmäßigkeit* der Rechnungslegung umfaßt die Prüfung der Einhaltung sowohl der gesetzlichen Ordnungsvorschriften als auch der unternehmensinternen Ordnungsvorschriften. Als ein wichtiger Sonderfall der unternehmensinternen Ordnungsvorschriften ist die Einhaltung des Systems der *Doppik* anzusehen, das wegen der doppelten Verbuchung jedes Geschäftsvorfalles hervorragend geeignet ist, Irrtümer oder unbefugte Veränderungen zu verhindern oder zu erschweren und dadurch die Sicherheit der Darstellung der Geschäftsvorfälle gegenüber den anderen Buchführungssystemen, wie der einfachen Buchführung oder der Kameralistik, zu erhöhen.

Durch die Prüfung der *materiellen Richtigkeit* der Rechnungslegung soll festgestellt werden, ob die Bestände und die Geschäftsvorfälle des Unternehmens gemäß den gesetzlichen Vorschriften bewertet und ausgewiesen werden.

Entsprechend dieser Einteilung werden anschließend die Prüfungsfragen nach der Vollständigkeit, der formellen Ordnungsmäßigkeit und der materiellen Richtigkeit der Rechnungslegung daraufhin untersucht, ob und in welchem Umfang sie mit Hilfe der zufallsgesteuerten Stichproben beantwortet werden können. Die Prüfung der Einhaltung der *Doppik* wird als besonders wichtiger Fall der Ordnungsmäßigkeitsprüfung gesondert behandelt werden.

¹⁾ s. § 135 Abs. 1 Satz 1 und Abs. 2 AktG.

II. Prüfung der Vollständigkeit der Rechnungslegung

1. Prüfung der Vollständigkeit der Bestandsaufnahme

Die Prüfung der mengenmäßigen Erfassung der *körperlichen Bestände* erfolgt

- a) durch persönliche Teilnahme des Prüfers an der Aufnahme der Bestände durch das Unternehmen oder
- b) nach der Aufnahme durch das Unternehmen an Hand der Inventurunterlagen, d. h. der Aufzeichnungen über die Aufnahme körperlicher Gegenstände in den Büchern und Schriften des Unternehmens.

Im Zuge dieser Prüfungen wird die *Vollständigkeit* des Prüfungsgegenstandes beobachtet.

Während die Aufdeckung von *Unterbeständen*, d. h. von fiktiven Inventarposten, im Rahmen der materiellen Prüfung erfolgt, weil das Unternehmen dem Prüfer grundsätzlich das Vorhandensein jedes durch einen Inventarposten ausgewiesenen Bestandspostens nachweisen muß, begegnet die Erfassung von *Überbeständen*, d. h. von tatsächlich vorhandenen, aber im Inventarverzeichnis nicht erfaßten Bestandsposten, besonderen Schwierigkeiten. Unter Umständen lagern Bestände im Zeitpunkt der Inventur an unbekannten Orten — z. B. auf betriebsfremdem Boden —, von deren Vorhandensein weder das Aufnahmepersonal noch der Prüfer Kenntnis zu haben brauchen, und diese Bestände sind möglicherweise auch von der Buchführung nicht erfaßt.

Aus den bei der Inventur erfaßten Bestandsposten kann aber weder durch bewußt ausgewählte Einheiten noch durch Zufallstichproben auf das Vorhandensein und auf die Größe von *nicht erfaßten Bestandsposten* geschlossen werden. Charakteristikum der Masse der tatsächlich vorhandenen Bestände ist in diesem Fall, daß ihr Umfang (die Zahl der Bestandsposten) für den Prüfer nicht bestimmbar ist. Selbst eine lückenlose Erfassung ist in diesem Fall mangels Kenntnis aller Daten über die fehlenden Bestände nicht durchführbar.

Da geeignete Prüfungsmethoden zur Feststellung von Überbeständen oftmals fehlen, schränkt der Prüfer seine Verantwortung für die Richtigkeit seiner Prüfungsergebnisse dadurch ein, daß er sich von der Unternehmensleitung eine *Vollständigkeitserklärung* geben läßt, in der bestätigt wird, daß alle tatsächlich vorhandenen Bestände im Inventarverzeichnis erfaßt sind, daß also die dem Prüfer bekanntgegebenen Bestände mit den wirklich vorhandenen übereinstimmen.

Die Glaubwürdigkeit der Vollständigkeitserklärung läßt sich durch Kontrollrechnungen und -überlegungen sowie Betriebs- und Lagerstättenbegehungen in gewissem Umfang nachprüfen. Aber selbst wenn der Prüfer hierbei im Inventarverzeichnis nicht erfaßte Bestandsposten entdeckt, kann er nur die unmittelbaren Korrekturen vornehmen; eine Schätzung von Zahl und Umfang oder Wert *aller* nicht erfaßten Posten ist auch mit statistischen Mitteln nicht möglich.

Soweit es sich um *nicht körperliche Bestände*, also um Forderungen, Verpflichtungen, Beteiligungen usw., handelt, ergeben sich die gleichen oben geschilderten Schwierigkeiten der prüferischen Erfassung unbekannter Bestandsposten. Auch hier kann nur die „*Vollständigkeitserklärung*“ den Verantwortungsbereich des Prüfers entsprechend einschränken.

2. Prüfung der Vollständigkeit der Erfassung der Geschäftsvorfälle

Die materielle Richtigkeit der Buchführung und des Jahresabschlusses erfordert, daß *alle buchungspflichtigen Geschäftsvorfälle* und nur diese erfaßt werden.

Die Frage, ob jeder *Buchung* auch ein Geschäftsvorfall entspricht, kann im Rahmen der materiellen Prüfung bei Anwendung des retrograden Prüfungsweges beantwortet werden.

Die Suche nach *nicht verbuchten Geschäftsvorfällen* begegnet den gleichen Schwierigkeiten wie die Suche nach nicht inventarisierten Beständen. Einmal können nicht alle überhaupt vorhandenen Prüfungsunterlagen („Handelsbriefe“ etc.) geprüft werden, um nicht verbuchte Geschäftsvorfälle zu finden. Eine Untersuchung dieser Art scheitert in aller Regel allein schon aus Zeit- und Kostengründen. Außerdem kann es Geschäftsvorfälle geben, die keinen schriftlichen Niederschlag gefunden haben oder deren schriftliche Unterlagen irrtümlich oder absichtlich verlegt worden oder abhanden gekommen sind.

Die Vollständigkeit der Erfassung aller buchungspflichtigen Geschäftsvorfälle läßt sich daher weder durch Vollerhebungen noch durch Teilerhebungen, und zwar weder durch bewußt ausgewählte noch durch Zufallsstichproben, nachprüfen. Werden nicht erfaßte buchungspflichtige Vorfälle durch glückliche Umstände aufgedeckt, so ist nur ihre Einzelberichtigung möglich und notwendig, doch kann von den dabei festgestellten Daten nicht auf weitere, vermutlich nicht erfaßte buchungspflichtige Geschäftsvorfälle geschlossen werden.

Die einzige Möglichkeit, die Verantwortung für eine etwaige Differenz zwischen der größeren wirklichen Zahl der Geschäftsvorfälle und der Zahl der verbuchten Geschäftsvorfälle abzugrenzen, ist für den Prüfer wiederum die „*Vollständigkeitserklärung*“ der Unternehmensleitung.

III. Prüfung der formellen Ordnungsmäßigkeit der Rechnungslegung

1. Prüfung der Einhaltung des Systems der doppelten Buchführung

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Im System der doppelten Buchführung (Doppik) erfolgt die Verbuchung jedes Geschäftsvorfalles mit dem gleichen Gesamtbetrag einmal in einem oder mehreren Konten im Soll und zum anderen in einem oder mehreren Konten im Haben. Daher ergibt sich nach vollständiger Verbuchung jedes Geschäftsvorfalles stets *Summengleichheit* zwischen allen Soll- und allen Habenbuchungen.

Ferner wird jeder Geschäftsvorfall einmal in *zeitlicher Reihenfolge* im Journal oder in Grundbüchern (im Soll und im Haben) und zum anderen nach *sachlichen Gesichtspunkten* gegliedert in den Sachkonten (im Soll und im Haben) verbucht.

Demgemäß sind zunächst folgende *Prüfungsfragen* zu beantworten:

- a) Besteht Summengleichheit zwischen allen Sollbuchungen und allen Habenbuchungen in den Journalen bzw. Grundbüchern?
- b) Besteht Summengleichheit zwischen allen Sollbuchungen und allen Habenbuchungen in der Gesamtheit der Sachkonten?
- c) Besteht Summengleichheit zwischen den Sollbuchungen und den Habenbuchungen in den Journalen bzw. Grundbüchern einerseits und den Sollbuchungen und den Habenbuchungen in den entsprechenden Sachkonten andererseits?

Ist in allen Fällen Summengleichheit gegeben, so besteht die *Vermutung*, aber noch nicht die Gewißheit, daß das System der Doppik eingehalten worden ist. Es ist nämlich möglich, daß Additions-, Übertragungs- oder Saldierungsfehler vorgekommen sind, die sich ganz oder teilweise gegenseitig aufheben. Daher muß im Rahmen der Systemprüfung weiter gefragt werden:

- d) Stimmen die Additionen auf den einzelnen Konten?
- e) Stimmen die Überträge innerhalb der einzelnen Konten (von Seite auf Seite) und von Konto auf Konto?
- f) Ist die Differenz zwischen den Summen der Sollbuchungen und der Habenbuchungen eines Kontos, also der Saldo, rechnerisch richtig ermittelt?

In allen Fällen sind auch *Pfennigdifferenzen* störend. Sie haben das gleiche Gewicht wie Differenzen größeren Ausmaßes, weil die einzelne Differenz der Saldo mehrerer Fehler sein kann. Mehrfache, sich gegenseitig aufhebende Fehler heilen also nicht die Systemverletzung, sondern bedeuten mehrfache Mängel, die erst aufgefunden und beseitigt werden müssen, ehe die weitere Prüfung auf dem System der Doppik mit der ihr eigenen Kontrollwirkung aufbauen kann.

Anders sind *Fehlbuchungen* zu beurteilen, also etwa Verbuchungen von Geschäftsvorfällen auf falschen Konten (z. B. Lohnzahlungen auf Konto „Gehälter“ statt auf Konto „Löhne“) oder mit falschen Beträgen (z. B. DM 1000,— statt DM 10 000,—). Solche Fehlbuchungen berühren das System der Doppik (Summengleichheit etc.) nicht. Sie beeinträchtigen aber die materielle Richtigkeit des Jahresabschlusses. Diese Fehler sind daher nicht im Rahmen der Systemprüfung, sondern später bei der Prüfung der materiellen Richtigkeit der einzelnen Geschäftsvorfälle zu berücksichtigen.

Wenn der Prüfer die mit der scharfen Kontrollfunktion der Doppik verbundene Sicherheit seinen weiteren Prüfungen zugrunde legen will, muß er also grundsätzlich *jeden Systemfehler* aufklären und berichtigen.

Absolute Sicherheit, d. h. die Wahrscheinlichkeit von 100%, für die Fehlerfreiheit der Doppik erreicht der Prüfer, wie in den folgenden Ausführungen dargelegt wird, nur durch eine *lückenlose Prüfung* aller Einheiten, jedoch niemals mit noch so umfangreichen Stichproben.

Ein *Abweichen vom Grundsatz der lückenlosen Prüfung* wird vom Prüfungsstoff aber dann erzwungen, wenn er so umfangreich ist, daß eine Vollerhebung aus zeitlichen oder aus Kostengründen nicht durchgeführt werden kann. In diesem

Fall ergibt sich die Frage, wie groß eine *Stichprobe* sein muß, damit eine *bestimmte Wahrscheinlichkeit* dafür erreicht wird, daß

- a) alle fehlerhaften Einheiten in die Stichprobe gelangen oder daß
- b) die Annahme, der nicht in die Stichprobe gezogene Teil der Grundgesamtheit sei fehlerfrei, richtig ist.

Diese Fragen lassen sich grundsätzlich mit Hilfe von „*Entdeckungsstichproben*“ beantworten.

b) Prüfung einer Grundgesamtheit auf Fehlerfreiheit mittels Entdeckungsstichproben

(1) Formelapparat

Um eine Antwort auf die Frage a) zu finden, nämlich wie groß der Stichprobenumfang sein muß, damit mit bestimmter Wahrscheinlichkeit alle in der Grundgesamtheit vorhandenen Fehler in die Stichprobe gelangen, sei angenommen, eine Grundgesamtheit vom Umfang N enthalte M fehlerhafte Einheiten; die übrigen $(N - M)$ Einheiten seien fehlerfrei. Dann ist die Wahrscheinlichkeit $Pr_{m; M}$, daß m der M Fehler in eine Stichprobe vom Umfang n gelangen und $(M - m)$ Fehler im nicht geprüften Rest vom Umfang $(N - n)$ der Grundgesamtheit verbleiben, gegeben durch

$$(1) \quad Pr_{m; M} = \frac{\binom{n}{m} \binom{N-n}{M-m}}{\binom{N}{M}} = \binom{M}{m} \frac{n!}{(n-m)!} \frac{(N-n)!}{[N-n-(M-m)]!} \frac{1}{N!} \frac{(N-M)!}{(N-m)!}$$

Wenn m sehr klein im Vergleich zu n und $(M - m)$ sehr klein gegenüber $(N - n)$ sind, was beim vorliegenden Prüfungsproblem im allgemeinen zutrifft, und wenn ferner $n \ll N$, kann man als Näherungswert für (1) die folgende Formel nehmen

$$(2) \quad Pr_{m; M} = \binom{M}{m} \left(\frac{n}{N}\right)^m \left(\frac{N-n}{N}\right)^{M-m}$$

Mit Hilfe dieser Formel kommt man einer Antwort auf die am Ende des vorigen Abschnitts gestellte Frage a) näher, indem zunächst (2) für $m = M$ gelöst wird. Man erhält dann

$$(3) \quad Pr_{m=M} = \left(\frac{n}{N}\right)^M$$

und die Auflösung von (3) nach n ergibt

$$(4) \quad n = N \sqrt[M]{Pr_{m=M}}$$

Dividiert man beide Seiten von (4) durch N , so sieht man, daß der erforderliche Auswahlatz der Entdeckungsstichprobe ausschließlich von der gewünschten Entdeckungswahrscheinlichkeit und der Fehlerzahl in der Grundgesamtheit,

nicht aber vom Umfang dieser Grundgesamtheit abhängt. Tabelle 1 enthält eine Zusammenstellung der unter bestimmten Bedingungen erforderlichen Stichprobenumfänge, berechnet nach Formel (4).

Tabelle 1: Erforderlicher Stichprobenumfang n absolut und in % von N für das Erreichen der vorgegebenen Wahrscheinlichkeit $Pr_{m=M}$ der Entdeckung aller vermutlich vorhandenen Fehler bei gegebenem Umfang der Grundgesamtheit N und gegebener vermutlicher Fehlerzahl M .

Umfang der Grundgesamtheit N	Fehlerzahl in der Grundgesamtheit M	Erforderlicher Stichprobenumfang für eine Entdeckungswahrscheinlichkeit $Pr_{m=M}$ von					
		60%		80%		90%	
		absolut	in % von N	absolut	in % von N	absolut	in % von N
2 000	1	1 200	60	1 600	80	1 800	90
	2	1 549	77	1 789	89	1 898	95
	3	1 687	84	1 857	93	1 931	97
10 000	1	6 000	60	8 000	80	9 000	90
	2	7 746	77	8 945	89	9 487	95
	3	8 435	84	9 283	93	9 655	97
100 000	1	60 000	60	80 000	80	90 000	90
	2	77 460	77	89 444	89	94 868	95
	3	84 343	84	92 830	93	96 550	97

Ebenso wie die Formel (4) läßt die Tabelle erkennen, daß bei größerer Fehlerzahl und größerer erforderlicher Entdeckungswahrscheinlichkeit ein größerer Auswahlatz benötigt wird.

Um z. B. von den vermutlich drei vorhandenen Fehlern alle in die Stichprobe zu bekommen, müßten nach der Tabelle unabhängig vom Umfang der Grundgesamtheit für eine Entdeckungswahrscheinlichkeit von 60% ein relativer Stichprobenumfang von rd. 84% der Grundgesamtheit, für eine Wahrscheinlichkeit von 80% ein Auswahlatz von rd. 93% der Grundgesamtheit und für eine Wahrscheinlichkeit von 90% eine Stichprobe von rd. 97% der Grundgesamtheit geprüft werden.

Nun kann aber die Zahl der Verstöße gegen das System der Doppik nicht vorhergesagt werden. Ereignisse, die das System verletzen, können, wie schon aus der Vielzahl der eingangs angeführten Prüfungsfragen hervorgeht, höchst unterschiedliche Ursachen haben. In einem Jahr kann es sich z. B. um Fehler in den Additionen oder Überträgen, im anderen um Unterschiede zwischen einzelnen korrespondierenden Soll- und Habenbuchungen, im nächsten Jahr um das Vorhandensein beider Arten handeln. Ein andermal liegt gar kein Fehler vor. Es ist daher praktisch nicht möglich, z. B. aus der Art und dem Umfang der in den Vorjahren aufgedeckten Systemverletzungen auf den Charakter und die Zahl der im zu prüfenden Geschäftsjahr vorgekommenen Systemverletzungen zu schließen.

Hieraus ergibt sich, daß mit der vorgenannten Formel (4) wegen der Unbestimmbarkeit der Zahl der in der Grundgesamtheit vorhandenen Fehler die eingangs gestellte Frage a) — vgl. S. 21 — nicht beantwortet werden kann.

Will man die *Frage b)* — vgl. S. 21 — beantworten, nämlich wie groß die Stichprobe sein muß, damit eine *bestimmte Wahrscheinlichkeit für die Fehlerfreiheit* des nicht in die Stichprobe gezogenen Teils der Grundgesamtheit erreicht wird, so kann man zunächst prüfen, ob die Hypothese, die Grundgesamtheit sei fehlerfrei, berechtigt ist. Die Hypothese soll angenommen werden, wenn *in der Stichprobe kein Fehler festgestellt* wird, andernfalls soll die Hypothese zurückzuweisen sein. Das Risiko, die Hypothese der Fehlerfreiheit zu verwerfen, obwohl sie richtig ist, d. h. obwohl keine Fehler in der Grundgesamtheit vorhanden sind, ist bei jedem Stichprobenumfang gleich Null (es kann auch in der Stichprobe kein Fehler auftreten, wenn die Grundgesamtheit keinen enthält). Das Risiko, die Hypothese anzunehmen, obwohl die Grundgesamtheit Fehler enthält, läßt sich berechnen: Die Wahrscheinlichkeit, keinen Fehler in einer Stichprobe vom Umfang n aus einer Grundgesamtheit vom Umfang N mit M Fehlern zu finden, ist gemäß Formel (1) bei $M = 1$ mit

$$(5) \quad Pr_{m=0; M=1} = \frac{N-n}{N}$$

am größten.

Umgekehrt ist die Wahrscheinlichkeit für die berechtigte Rückweisung der Hypothese, die Grundgesamtheit sei fehlerfrei, am kleinsten, wenn $m = M = 1$, nämlich

$$(6) \quad Pr_{m=M=1} = 1 - Pr_{m=0; M=1} = \frac{n}{N}$$

Soll also eine bestimmte Wahrscheinlichkeit für die Fehlerfreiheit der Grundgesamtheit erreicht werden, so ist $Pr_{m=M=1}$ gleich dieser Wahrscheinlichkeit zu setzen und der Stichprobenumfang so zu bemessen, daß

$$(7) \quad n = N Pr_{m=M=1}$$

Kann eine Annahme darüber gemacht werden, welche A-priori-Wahrscheinlichkeiten, d. h. welche Wahrscheinlichkeiten vor der Erhebung für das Vorhandensein von K Fehlern ($K = 0, 1, 2, \dots, N$), in der Grundgesamtheit bestehen, so ist ein anderes Vorgehen möglich:

Sind in der Stichprobe m Fehler gefunden worden und kann unterstellt werden, daß die A-priori-Wahrscheinlichkeiten für das Vorhandensein von K Fehlern gleich sind¹⁾, so beträgt die Wahrscheinlichkeit für das Vorhandensein von k Fehlern ($k = 0, 1, \dots, N - n$) im nicht geprüften Rest der Grundgesamtheit aufgrund des *Bayesschen* Rückschlusses

$$(8) \quad Pr_{M \leq m+k} = \frac{\sum_{K=m}^{m+k} \binom{K}{m} \left(\frac{n}{N}\right)^m \left(\frac{N-n}{N}\right)^{K-m}}{\sum_{K=m}^{m+N-n} \binom{K}{m} \left(\frac{n}{N}\right)^m \left(\frac{N-n}{N}\right)^{K-m}}$$

¹⁾ Siehe hierzu *Coulinet, J.*, Application de la technique des sondages à la revision comptable, Rapport de synthèse der Verhandlungen der Kommission Mathématiques Financières der UEC, Nizza 1958 (unveröffentlichtes Manuskript), S. 6 ff.

Setzt man $k = 0$, so erhält man

$$(9) \quad Pr_{M=m} = \frac{\left(\frac{n}{N}\right)^m}{\sum_{K=m}^{m+N-n} \binom{K}{m} \left(\frac{n}{N}\right)^m \left(\frac{N-n}{N}\right)^{K-m}}$$

Dies ist unter den vorstehenden Bedingungen die Wahrscheinlichkeit dafür, daß im Rest der Grundgesamtheit kein Fehler mehr vorhanden ist. Eine Auflösung dieser Formel nach n ist im allgemeinen Fall nicht möglich. Hiermit ist dann die eingangs (S. 21) gestellte Frage b) beantwortet.

(2) Wirkung der Bestimmungsfaktoren auf den Stichprobenumfang

Der Stichprobenumfang muß, wie (7) zeigt, mindestens das N -fache der erforderlichen Wahrscheinlichkeit für die Fehlerfreiheit der Grundgesamtheit betragen; m. a. W. der Auswahlatz muß mindestens die gleiche Größe haben wie diese Wahrscheinlichkeit.

Um also z. B. mit einer Sicherheit von 60% gegebenenfalls behaupten zu können, die Grundgesamtheit sei fehlerfrei, sind mindestens 60% der Einheiten der Grundgesamtheit zu prüfen. Werden 80% oder 90% Sicherheit verlangt, so sind 80% respektive 90% der Untersuchungseinheiten in die Stichprobe zu ziehen.

Taucht aber in der Stichprobe auch nur *ein* Fehler auf, so ist wie auf S. 23 beschrieben vorzugehen. Die Wahrscheinlichkeit, daß *weitere* Fehler vorhanden sind, vorausgesetzt, die A-priori-Wahrscheinlichkeiten für ihr Vorhandensein sind gleich, ergibt sich gemäß Formel (8). Sie ist größer, als wenn kein Fehler gefunden worden wäre. In diesem Fall muß also der Prüfer — der grundsätzlich Fehlerfreiheit verlangt — nicht nur den bzw. die gefundenen Fehler berichtigen, sondern den Prüfungsumfang ausdehnen, um sich ein Urteil über die Fehlerfreiheit des bislang nicht geprüften Restes der Grundgesamtheit zu verschaffen. Das bedeutet aber, wie auch Tabelle 1 zeigt, daß der Stichprobenumfang sehr groß werden muß, wenn eine hohe Sicherheit für die Fehlerfreiheit erreicht werden soll.

Umgekehrt folgt daraus, daß kleine Stichproben wegen der unbeachtlichen Wahrscheinlichkeiten für die Entdeckung vermuteter seltener Fehler reine Verschwendung sind und ebenso gut unterbleiben können¹⁾.

c) Anwendbarkeit auf das Prüfungsgebiet

Kann also die eingangs gestellte Frage a), ob eine bestimmte Wahrscheinlichkeit dafür erreichbar ist, daß alle fehlerbehafteten Einheiten in die Stichprobe gelangen, mit Hilfe der Stichprobentheorie *nicht* gelöst werden, weil die Zahl der vorgekommenen Verstöße gegen die Doppik nicht angegeben werden kann, so erlaubt die Formel (6) eine Aussage über die Sicherheit für die Richtigkeit der Vermutung der Fehlerfreiheit der Grundgesamtheit, falls keine Fehlereinheit in der Stichprobe auftaucht (Frage b). Der notwendige Auswahlatz der Stichprobe ist in diesem Fall gleich der geforderten Sicherheit für die Fehlerfreiheit (Formel 7).

¹⁾ Arkin (1961), S. 53.

Bei der Abschlußprüfung ist aber nur die Feststellung, daß eine hohe Sicherheit für die Fehlerfreiheit des Systems der Doppik besteht, von Bedeutung. Grundsätzlich benötigt der Prüfer absolute, d. h. 100%ige, Sicherheit, die nach den vorstehenden Darlegungen nur durch eine *Vollerhebung* zu erreichen ist.

Da jedoch Vollerhebungen aus Zeit- und Kostengründen nur bei kleinen Buchführungen, in der Regel aber nicht durchführbar sind, teilt man in der Prüfungspraxis die Grundgesamtheit in Teilgesamtheiten, sogenannte *Fehlerfelder*, auf und prüft die wichtigsten von ihnen lückenlos.

Aber auch die lückenlose Prüfung der einzelnen Fehlerfelder ergibt immer nur die Sicherheit über die Fehlerfreiheit innerhalb dieser Fehlerfelder und erlaubt keinen Wahrscheinlichkeitsschluß auf die Fehlerfreiheit der übrigen nicht geprüften Fehlerfelder. Zweckmäßigerweise sind dann diejenigen zur lückenlosen Prüfung heranzuziehen, innerhalb derer das innerbetriebliche Kontrollsystem nicht ausreichend funktioniert. Auf die Prüfung der Wirksamkeit des innerbetrieblichen Kontrollsystems wird im übernächsten Abschnitt eingegangen werden.

2. Prüfung der Einhaltung der sonstigen formellen Ordnungsvorschriften

A. Prüfung der Einhaltung der gesetzlichen Ordnungsvorschriften

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Die *formellen Ordnungsvorschriften* sollen — wie eingangs dargestellt — die Verständlichkeit und Sicherheit der Rechnungslegung gewährleisten. Diese Vorschriften sind teilweise vom *Gesetzgeber*, teilweise durch die *Unternehmen* selbst erlassen worden.

Ziel der Prüfungen ist die Feststellung, ob und in welchem Umfang diese Vorschriften eingehalten bzw. verletzt worden sind; z. B. ob und in welchem Umfang das Verbot von Rasuren beachtet, leere Zwischenräume zwischen den Eintragungen entwertet, die Eintragungen in der richtigen Reihenfolge vorgenommen, die Landeswährung für die Wertangaben benutzt, eine lebende Sprache gebraucht worden sind u. a. m.

Von diesen Vorschriften sind diejenigen, die die *Sicherheit* der Rechnungslegung betreffen, die bei weitem wichtigsten, während Verstöße gegen die Vorschriften zur Sicherung der *Verständlichkeit* der Rechnungslegung (z. B. Gebrauch einer lebenden Sprache) praktisch kaum vorkommen.

Zwischen den *gesetzlichen* und den *unternehmensinternen* Ordnungsvorschriften besteht prüferisch ein Unterschied insofern, als Verstöße gegen die gesetzlichen Vorschriften zur Verweigerung oder Einschränkung des Bestätigungsvermerks führen können, Verstöße gegen die unternehmensinternen Ordnungsvorschriften dagegen keine solchen Folgen haben.

Bei der Prüfung der Einhaltung der *gesetzlichen* Vorschriften muß der Prüfer also nicht ausschließlich den Fehleranteil feststellen, sondern ferner entscheiden, ob der Umfang der Verstöße gegen sie so groß ist, daß er den Bestätigungsvermerk verweigern oder einschränken muß. Will der Prüfer nur diese letzte Entscheidung

fallen, genügt ihm die Erkenntnis, ob eine Grundgesamtheit mit bestimmter Wahrscheinlichkeit *weniger als einen bestimmten Prozentsatz* Fehler enthält. Verfahren, die diese Entscheidung auf Stichprobenbasis in rationeller Weise erlauben, sind die „*Annahmestichproben*“.

b) *Entscheidung über Annahme oder Rückweisung einer Grundgesamtheit mittels Annahmestichproben*

(1) *Formelapparat*

Maßgeblich für die Güte einer Annahmestichprobe ist die bei einem bestimmten Prüfplan gegebene Wahrscheinlichkeit, eine Grundgesamtheit mit geringem Fehleranteil anzunehmen bzw. eine Grundgesamtheit mit hohem Fehleranteil zurückzuweisen.

Bezeichnet man den Umfang der Grundgesamtheit mit N , die Fehlerzahl in der Grundgesamtheit mit M , den Stichprobenumfang mit n und die Zahl der fehlerhaften Stichprobeneinheiten mit m , so ist die *Annahmewahrscheinlichkeit*, d. h. die Wahrscheinlichkeit dafür, daß die Fehlerzahl in der Stichprobe die Annahmehzahl a nicht überschreitet, gegeben durch

$$(10) \quad Pr_{M; m \leq a} = \sum_{m=0}^a \frac{\binom{n}{m} \binom{N-n}{M-m}}{\binom{N}{M}} = \sum_{m=0}^a \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$$

Als gute Näherungswerte für $Pr_{M; m \leq a}$ gelten in der Stichprobentheorie¹⁾

a) wenn die Grundgesamtheit im Vergleich zur Stichprobe sehr groß ist,

$$(11) \quad Pr_{\frac{M}{N}; n; m \leq a} = \sum_{m=0}^a \binom{n}{m} \left(\frac{M}{N}\right)^m \left(\frac{N-M}{N}\right)^{n-m}$$

b) wenn die Grundgesamtheit im Vergleich zur Zahl der in ihr enthaltenen Fehler sehr groß ist (vgl. S. 21)

$$(12) \quad Pr_{\frac{M}{N}; \frac{n}{N}; m \leq a} = \sum_{m=0}^a \binom{M}{m} \left(\frac{n}{N}\right)^m \left(\frac{N-n}{N}\right)^{M-m}$$

c) wenn sowohl $M/N < 0,10$ als auch $n/N < 0,10$ sind

$$(13) \quad Pr_{\frac{n}{N}; \frac{M}{N}; m \leq a} = \sum_{m=0}^a \frac{1}{m!} \left(n \frac{M}{N}\right)^m e^{-n \frac{M}{N}}$$

wobei $e = 2,71828 \dots$ die Basis des natürlichen Logarithmus ist.

¹⁾ So z. B. auch Dodge, H. F./Romig, H. G., Sampling Inspection Tables, 3. Aufl., New York und London 1946, S. 43, 44.

Umgekehrt beträgt die Wahrscheinlichkeit, daß in dieser Stichprobe mindestens r (Rückweisungszahl) Fehler auftauchen

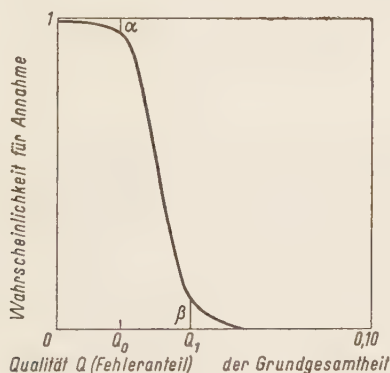
$$(14) \quad Pr_{M; m \geq r} = 1 - \sum_{m=0}^{r-1} \frac{\binom{n}{m} \binom{N-n}{M-m}}{\binom{N}{M}} = 1 - \sum_{m=0}^{r-1} \frac{\binom{M}{m} \binom{N-M}{n-m}}{\binom{N}{n}}$$

Sie wird als „Rückweisungswahrscheinlichkeit“ bezeichnet.

Näherungswerte hierfür lassen sich ebenso wie für die Formel (10) bilden bzw. verwenden.

Die Annahme- und Rückweisungswahrscheinlichkeiten für eine Grundgesamtheit mit bestimmtem Fehleranteil ergänzen sich zum Wert Eins, sofern $r = a + 1$.

Die Abhängigkeit der Wahrscheinlichkeiten (10) bzw. (14) und ihrer Gegenwahrscheinlichkeiten vom Fehleranteil in der Grundgesamtheit bei gegebenem n und a bzw. r läßt sich graphisch darstellen. Figur 1 ist ein Beispiel dafür. Man bezeichnet diese Kurve als *Prüfplankurve* oder in Anlehnung an den angelsächsischen Ausdruck „operating characteristic curve“ als „Operations-Charakteristik“ bzw. kurz als „OC-Kurve“.



Figur 1: Prüfplankurve für $N = 100\,000$; $n = 582$; $a = 17$; $r = 18$

Der Verlauf einer solchen Kurve ist durch zwei Punkte eindeutig bestimmt, sofern die Relation zwischen a und r vorgegeben ist, z. B. wenn $r = a + 1$. Bezeichnet man den durchschnittlichen Fehleranteil mit Q_0 , das Rückweisungsrisiko mit α , die Rückweisungsgrenze mit Q_1 und das für sie akzeptierte Annahmerisiko mit β , so ist die Prüfplankurve durch Vorgabe der Punkte $(Q_0; 1 - \alpha)$ und $(Q_1; \beta)$ festgelegt, z. B. in Figur 1 durch $(Q_0 = 0,02; 1 - \alpha = 0,95)$ und $(Q_1 = 0,04; \beta = 0,10)$.

Der gewünschte Verlauf der Prüfplankurve kann durch *ein- und mehrphasige Stichprobenverfahren* sowie durch *Sequenzstichproben* erreicht werden.

Bei *einphasigen* Stichproben gilt die Beziehung $a = r - 1$. *Mehrphasige Verfahren* zeichnen sich dadurch aus, daß in allen Phasen mit Ausnahme der letzten, für die die vorgenannte Beziehung gilt, zwischen den Annahme- und Rückweisungszahlen die Beziehung $a_i < r_i - 1$ besteht, wobei i die Ordnungszahl der Phase ist und folglich noch eine *dritte Entscheidungsmöglichkeit*, nämlich die *Vergrößerung des Stichprobenumfangs* durch Fortsetzung der Auswahl in einer weiteren Phase ($i + 1$) gegeben ist. Die Stärke der Einengung des Spielraums zwischen a_i und $(r_i - 1)$ von Phase zu Phase richtet sich bei gleichen sonstigen Voraussetzungen nach der vorgesehenen Zahl der Auswahlphasen.

Die Verallgemeinerung des *Formelapparates* auf *mehrphasige Stichproben* bietet keine prinzipiellen Schwierigkeiten, wenngleich dieser Apparat wesentlich komplizierter ist als für die einphasigen Annahmestichproben. Letzteres ist insbesondere deshalb der Fall, weil die vorgeschriebenen Annahme- und Rückweisungsrisiken angemessen auf die einzelnen Phasen verteilt werden müssen¹⁾.

Die einer zweiphasigen Auswahl zugehörige Formel für die Annahmewahrscheinlichkeit β lautet z. B.

$$\begin{aligned}
 \beta = & \sum_{m=0}^{a_1} Pr_{m; n_1; N; M} + Pr_{a_1+1; n_1; N; M} + \sum_{m=0}^{a_2-a_1-1} Pr_{m; n_2; N-n_1; M-a_1-1}; \\
 (15) \quad & + Pr_{a_1+2; n_1; N; M} + \sum_{m=0}^{a_2-a_1-2} Pr_{m; n_2; N-n_1; M-a_1-2}; \\
 & + \dots \\
 & + Pr_{a_2; n_1; N; M} + Pr_{0; n_2; N-n_1; M-a_2}
 \end{aligned}$$

Dabei sind N der Umfang der Grundgesamtheit, M die Fehlerzahl in der Grundgesamtheit, n_1 der Stichprobenumfang der ersten Phase, a_1 die Annahmezah der ersten Phase, n_2 der Stichprobenumfang für die zweite Phase und a_2 die kumulierte Annahmezah der ersten und zweiten Phase, d. h. die für die Annahme zulässige Gesamtzahl der in der ersten und zweiten Phase festgestellten fehlerhaften Stichprobeneinheiten.

Somit ist β die Wahrscheinlichkeit, in der Stichprobe der ersten Phase höchstens a_1 Fehler zu finden, vermehrt um die Wahrscheinlichkeiten, ($a_1 + 1$) Fehler in der ersten Phase und höchstens $(a_2 - a_1 - 1)$ Fehler in der zweiten Phase, ($a_1 + 2$) Fehler in der ersten Phase und höchstens $(a_2 - a_1 - 2)$ Fehler in der zweiten Phase zu finden usw., bis zur Wahrscheinlichkeit, in der ersten Phase a_2 Fehler und in der zweiten Phase keine fehlerhaften Stichprobeneinheiten festzustellen.

Bei den *Sequenzstichproben* erfolgt bei wachsendem Stichprobenumfang keine Einengung des Spielraums zwischen a_i und r_i . Das hat zur Folge, daß auch die Höchstzahl der Phasen (= Stichprobeneinheiten) nicht festgelegt werden kann.

¹⁾ Siehe hierzu *Dodge/Romig*, a. a. O., S. 43, 47.

Die für einen bestimmten Stichprobenumfang geltenden Annahme- und Rückweisungszahlen werden bei Sequenzstichproben nach folgenden Formeln¹⁾ berechnet:

$$(16) \quad a = \frac{\log \frac{\beta}{1-\alpha}}{\log \frac{Q_1}{Q_0} - \log \frac{1-Q_1}{1-Q_0}} + n \frac{\log \frac{1-Q_0}{1-Q_1}}{\log \frac{Q_1}{Q_0} - \log \frac{1-Q_1}{1-Q_0}}$$

$$r = \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{Q_1}{Q_0} - \log \frac{1-Q_1}{1-Q_0}} + n \frac{\log \frac{1-Q_0}{1-Q_1}}{\log \frac{Q_1}{Q_0} - \log \frac{1-Q_1}{1-Q_0}}$$

Die *Ermittlung der Stichprobenkriterien* (Umfang der Stichprobe bzw. Zahl und Umfang der Phasen, Annahme- und Rückweisungszahlen für die Stichprobe bzw. die einzelnen Phasen) ist, abgesehen von den Sequenzstichproben, nur mittels Herantasten durch wiederholte Berechnungen mit schrittweise verbesserten Näherungswerten möglich²⁾. Diese Schwierigkeiten haben dazu geführt, daß *Tabellen und Kurvendiagramme* erarbeitet wurden, aus denen die Kriterien für häufig vorkommende Bedingungen abgelesen werden können.

(2) Wirkung der Bestimmungsfaktoren auf den Stichprobenumfang

Mit der Anwendung von Annahmestichprobenverfahren sowohl bei der internen wie bei der externen Prüfung buchhalterischer Tatbestände haben sich insbesondere *Vance/Neter* in ihrem schon mehrfach zitierten Buch sehr eingehend befaßt. Sie geben dort auch Auszüge der wichtigsten Tabellen und Diagramme für Prüfpläne wieder und erläutern ihre Anwendung³⁾.

Für die Zwecke der Abschlußprüfung sind jedoch nicht alle der dort erwähnten Tabellen und Diagramme gleichermaßen geeignet. Die *Innenrevision* hat durch Prüfung der Rechnungslegung kurzfristiger Zeitabschnitte dafür zu sorgen, daß auf die Dauer ein bestimmtes Qualitätsniveau nicht unterschritten wird. Dagegen hat der *Abschlußprüfer*, wie eingangs dargelegt, die Aufgabe, die Rechnungslegung eines bestimmten Geschäftsjahres zu prüfen.

Bei der Abschlußprüfung sind daher nur solche Tabellen und Diagramme für Prüfpläne geeignet, mit denen man feststellen kann, ob mit bestimmter Sicherheit die Annahme berechtigt ist, daß die Rückweisungsgrenze Q_1 in der einzelnen zu prüfenden Grundgesamtheit nicht überschritten ist. Hierzu zählen z. B. die

¹⁾ *Coulinet, J.*, a. a. O., S. 6 ff.; s. auch *Vance/Neter*, a. a. O., S. 109 ff. und die dazu zitierte Literatur, insbesondere *Wald, A.*, *Sequential Analysis*, New York 1947.

²⁾ Eine Beschreibung des hierfür erforderlichen Vorgehens findet sich z. B. bei *Dodge/Romig*, a. a. O., S. 35, 45.

³⁾ a. a. O., S. 95 ff.

Military-Standard-105-A-Tabellen des US-Department of Defense, ferner diejenigen der *Dodge/Romig-Tabellen*, die für Rückweisungsgrenzen für einzelne Grundgesamtheiten (Partien) berechnet sind (LTPD-Tabellen), und die *Cameron-Tabelle*¹⁾.

Die einzelnen Tabellen sind für jeweils *andere Voraussetzungen* entwickelt worden. Der Prüfer muß daher die Risiken kennen, die in dem von einer bestimmten Tabelle gelieferten Prüfplan verkörpert sind, bevor er diesen Plan benutzt²⁾.

So ist z. B. das Rückweisungsrisiko für den jeweiligen durchschnittlichen Fehleranteil in den *Dodge/Romig-Tabellen* von Prüfplan zu Prüfplan verschieden. Jeder der Prüfpläne ist so berechnet, daß bei wiederholter Anwendung auf Grundgesamtheiten gleichen Umfangs und gleicher Struktur die Prüfungs- und Berichtigungsarbeiten einschließlich der lückenlosen Berichtigung des Prüfungsstoffs in den Fällen, wo die Stichprobe zur Ablehnung der Grundgesamtheit führte, im Durchschnitt je Anwendungsfall ein Minimum erreichen.

Bei der Anwendung dieser Tabellen muß man also wissen, daß das Rückweisungsrisiko nicht bei allen Plänen absolut gleich ist. Derartige Unterschiede in den Voraussetzungen sind neben den offenkundigen Unterschieden zwischen den Tabellen, wie z. B. durch Abweichungen in den Klassengrenzen, die Ursache für die häufige praktische Unvergleichbarkeit der verschiedenen Tabellen (vgl. z. B. S. 33, Fußnote 1).

Im folgenden soll aus den oben genannten Gründen nicht unter Zugrundelegung der Formeln (10) ff., sondern an Hand einiger der vorgenannten Tabellen der Einfluß

- a) des Umfangs der Grundgesamtheit,
- b) des Fehleranteils der Grundgesamtheit und seines Verhältnisses zur Rückweisungsgrenze,
- c) der Annahme- und Rückweisungsrisiken,
- d) der Art des Auswahlverfahrens,

auf den *Stichprobenumfang* näher dargestellt werden.

Die Abhängigkeit des Stichprobenumfangs vom *Umfang der Grundgesamtheit* sei zunächst an Beispielen gezeigt, die den *Dodge/Romig-Tabellen* entnommen sind.

In Tabelle 2 sind für bestimmte Größenklassen der Grundgesamtheit die bei ein- und zweiphasigen Auswahlverfahren erforderlichen absoluten Stichprobenumfänge zusammengestellt und ihr Verhältnis zum Umfang der Grundgesamtheit (Auswahlsatz in %) angegeben, die bei einer Rückweisungsgrenze $Q_1 = 0,04$, einem Annahmerisiko $\beta = 0,10$ sowie einem durchschnittlichen Fehleranteil in den Grenzen $0,0041 \leq Q_0 \leq 0,0080$ und einer Minimierung des durchschnittlichen Prüfungsumfangs je Partie (s. oben) erforderlich sind.

¹⁾ Munitions Board Standards Agency, Department of Defense, Sampling Procedures and Tables for Inspection by Attributes, Mil-Std-105 A, Government Printing Office, Washington DC 1950; *Dodge/Romig*, a. a. O., Table I, Single Sampling Lot Tolerance Tables, S. 70 ff., und Table II, Double Sampling Lot Tolerance Tables, S. 74 ff.; *Cameron, J. M.*, Tables for Constructing and for Computing the Operation Characteristics of Single Sampling Plans; Industrial Quality Control, Juli 1952, S. 37 ff.

²⁾ *Vance/Neter*, a. a. O., S. 89.

Tabelle 2: Absoluter und relativer Stichprobenumfang in Abhängigkeit vom Umfang der Grundgesamtheit bei ein- und zweiphasigen Prüfplänen für die Bedingungen $0,0041 \leq Q_0 \leq 0,0080$; $Q_1 = 0,04$; $\beta = 0,10$ nach Dodge/Romig-Tabellen¹⁾.

Umfang der Grundgesamtheit <i>N</i>	Stichprobenumfang bei				
	einphasiger Auswahl		zweiphasiger Auswahl		
	absolut	in % von <i>N</i>	Phase	absolut*)	in % von <i>N</i>
501— 600	125	25,0—20,8	I	65	13,0—10,8
			II	165	33,0—27,5
1 001— 2 000	165	16,5— 8,3	I	70	7,0— 3,5
			II	215	21,5—10,8
10 001— 20 000	265	2,7— 1,3	I	115	1,2— 0,6
			II	350	3,5— 1,8
50 001—100 000	325	0,7— 0,3	I	115	0,2— 0,1
			II	415	0,8— 0,4

*) Zweite Phasen: kumulierte Werte.

Die Tabelle zeigt, daß bei den Annahmestichproben, und zwar sowohl bei einphasigen wie bei zweiphasigen, mit *größerem Umfang der Grundgesamtheit ein größerer absoluter Stichprobenumfang* benötigt wird, *seine Zunahme erfolgt aber unterproportional zum Wachstum des Umfangs der Grundgesamtheit*. Würde man diese Entwicklung graphisch darstellen, so erhielte man für die einzelnen Phasen und für die kumulierten Umfänge jeweils ein ähnliches Bild, wie es die Figur 3 zeigt (s. S. 45).

Diese Beobachtung läßt sich in allen für bestimmte Rückweisungsgrenzen einzelner Grundgesamtheiten berechneten *Dodge/Romig-Tabellen* (LTPD-Tabellen) machen. Zu ähnlichen Beobachtungsergebnissen gelangt man auch bei den *Military-Standard-105-A-Tabellen*²⁾.

Die *Cameron-Tabelle* beruht auf der Verwandtschaft zwischen der *Poisson-Verteilung*³⁾ und einem bestimmten Kriterium für die Beurteilung, welche Höchstzahl von Fehlern in der Stichprobe als zufällig gelten kann, wenn bestimmte Bedingungen bezüglich der oben bezeichneten Punkte (Q_0 ; $1 - \alpha$) und (Q_1 ; β) der OC-Kurve eingehalten sind und welche Mindestzahl von Fehlern in der Stichprobe darauf schließen läßt, daß ein wesentlicher Unterschied zwischen den annehmbaren und den in der Grundgesamtheit tatsächlich herrschenden Gegebenheiten angenommen werden muß.

Benutzt man die *Cameron-Tabelle* zur Entwicklung eines Prüfplans, der die Bedingungen $Q_0 = 0,005$; $\alpha = 0,01$; $Q_1 = 0,04$; $\beta = 0,10$ erfüllt, so erhält man als erforderlichen Stichprobenumfang⁴⁾ $n = 256$ Einheiten.

1) Werte der Spalten 2 resp. 5 s. *Dodge/Romig*, a. a. O., und *Vance/Neter*, a. a. O., S. 82 resp. 84 f.; Werte der Spalten 3 und 6 ergeben sich aus den Werten der Spalten 1 und 2 resp. 5.

2) Vgl. *Vance/Neter*, a. a. O., S. 56.

3) Die *Poisson-Verteilung* gilt häufig für Merkmale, die nur ganzzahlige Werte annehmen können. Sie spielt insbesondere als Näherungsverteilung bei sehr kleinen Anteilen von Einheiten mit solchen Merkmalen in der Grundgesamtheit eine Rolle. Die Summenformel für eine *Poisson-Verteilung* ist z. B. Formel (13), S. 26.

4) *Cameron*, a. a. O., S. 39; *Vance/Neter*, a. a. O., S. 93 und S. 95, Abs. 2.

Nach der Faustregel für die Anwendung der *Poissonschen* Näherungsbeziehung, die besagt, daß die Bedingung $n < 0,10 N$ eingehalten sein muß¹⁾, kann dieser Prüfplan nur bei Grundgesamtheiten mit mindestens $N = 2560 \approx 3000$ Einheiten angewandt werden.

Aus allen Beispielen ergibt sich somit übereinstimmend, daß bei *Annahmestichproben* der Stichprobenumfang sehr stark seiner relativen Größe nach, verhältnismäßig wenig aber seiner absoluten Größe nach, vom Umfang der Grundgesamtheit beeinflusst wird, und zwar um so weniger, je größer der Umfang der Grundgesamtheit ist. Dieser zeigt damit hier eine gänzlich andere Wirkung wie bei den Entdeckungstichproben.

Auch für die Untersuchung der *Abhängigkeit des Stichprobenumfangs vom Fehleranteil der Grundgesamtheit* bedient man sich aus den auf Seite 29 genannten Gründen am besten der bereits erstellten und für die Abschlußprüfung in Frage kommenden Tabellen. In der folgenden Tabelle 3 sind die gemäß den schon erwähnten LTPD-Tabellen von *Dodge/Romig* erforderlichen Stichprobenumfänge für die in der Vorspalte genannten Größen von Q_1 und Q_0 bei ein- und zweiphasigen Prüfplänen jeweils für eine sehr große und eine verhältnismäßig kleine Grundgesamtheit zusammengestellt. Für die zweiphasigen Verfahren wird dabei in I die Zahl der in der ersten Phase auszuwählenden Einheiten genannt. In II ist die Zahl der in der ersten und zweiten Phase insgesamt auszuwählenden Einheiten angegeben, die in die Stichprobe zu ziehen sind, wenn in der ersten Phase die Zahl der fehlerhaften Stichprobeneinheiten größer als die Annahmezah und kleiner als die Rückweisungszahl war. Der durchschnittliche Fehleranteil und die Rückweisungsgrenze wurden dabei so festgelegt, daß in den Zeilen 1 und 2 die Differenz ($Q_1 - Q_0$) ungefähr gleich ist ($\approx 0,025$) und in den Zeilen 2 und 3 das Verhältnis (Q_1/Q_0) annähernd übereinstimmt ($\approx 2:1$). Im übrigen gelten die für die *Dodge/Romig-Tabellen* typischen Bedingungen, von denen auf Seite 30 die Rede war.

Tabelle 3: Stichprobenumfänge ein- und zweiphasiger Prüfpläne für Grundgesamtheiten bestimmter Größenordnung N in Abhängigkeit von bestimmten Größenverhältnissen der Rückweisungsgrenze Q_1 zum durchschnittlichen Fehleranteil Q_0 der Grundgesamtheit nach Dodge/Romig-Tabellen²⁾.

Q_1/Q_0	Einphasige Prüfpläne		Zweiphasige Prüfpläne		
	N		$N^*)$		
	501—600	50 001—100 000	Phase	501—600	50 001—100 000
0,03/(0,0031—0,0060)	120	390	I	85	150
			II	180	555
0,05/(0,0201—0,0250)	175	770	I	85	330
			II	250	1 140
0,03/(0,0121—0,0150)	200	1 215	I	135	515
			II	320	1 830

*) Zweite Phasen: kumulierte Werte.

1) Vgl. S. 26.

2) a. a. O.

Liegt demnach in einer Grundgesamtheit von 501 bis 600 Einheiten der durchschnittliche Fehleranteil zwischen $3,10/00$ und $6,00/00$ und ist die Rückweisungsgrenze auf 3% festgesetzt, so ist nach den *Dodge/Romig-Tabellen* bei einem einphasigen Prüfplan ein Stichprobenumfang von 120 Einheiten, bei einem zweiphasigen Prüfplan in der ersten Phase ein Stichprobenumfang von 85 und gegebenenfalls in der zweiten Phase ein solcher von 95 Einheiten erforderlich ($85 + 95 = 180$). Liegt die durchschnittliche Qualität einer gleichumfangreichen Grundgesamtheit aber zwischen 1,2% und 1,5% und ist als Rückweisungsgrenze ein Fehleranteil von 3% bestimmt, so muß der Stichprobenumfang bei einphasigem Prüfplan 200 Einheiten und bei zweiphasigem Prüfplan in der ersten Phase 135 sowie gegebenenfalls in der zweiten Phase weitere 185 Einheiten betragen ($135 + 185 = 320$).

Allgemein gilt, daß der *Stichprobenumfang bei gegebenen Bedingungen für Q_1 , α und β derart vom Fehleranteil in der Grundgesamtheit abhängt, daß ein kleinerer durchschnittlicher Fehleranteil Q_0 kleinere Stichprobenumfänge, d. h. schnellere Entscheidungen über Annahme oder Rückweisung der Grundgesamtheit, erlaubt*. Die Ursache liegt darin, daß, je kleiner Q_0 , desto größer die Differenz $Q_1 - Q_0$ wird.

Wie ferner zu erkennen ist, hängt der *Umfang der Stichprobe* nicht nur von der Größe des kritischen Bereichs, sondern *auch von seinem Abstand zum Punkt völliger Fehlerfreiheit ($Q = 0$)* ab, m. a. W. von seiner relativen Größe in bezug auf den durchschnittlichen Fehleranteil Q_0 . *Je geringer die Fehleranteile werden, desto kleiner kann bei gleicher absoluter Größe des kritischen Bereichs der Umfang der Stichprobe sein und umgekehrt*.

Diese Ergebnisse werden durch die Zahlenzusammenstellungen in Tabelle 4, die auf der *Cameron-Tabelle* beruhen, bestätigt¹⁾.

Tabelle 4: Stichprobenumfang in Abhängigkeit von bestimmten Qualitäts- und Risikobedingungen gemäß Cameron-Tabelle²⁾

Q_1/Q_0	Risikobedingungen α/β			
	0,01/0,10	0,01/0,05	0,05/0,01	0,01/0,01
0,03/0,005	256	256	274	466
0,05/0,025	719	909	932	1 269
0,03/0,015	1 198	1 515	1 554	2 114

Darüber hinaus lassen sich aus Tabelle 4 einige Beispiele für die *Wirkung der Veränderung der Risikobedingungen auf den Stichprobenumfang* bei gegebenen Qualitätsbedingungen entnehmen: Mit einer Ausnahme steigen in allen Zeilen von der linken zur rechten Spalte fortschreitend die Stichprobenumfänge an. Die Ausnahme findet sich in den ersten beiden Spalten der ersten Zeile, wo die

¹⁾ Obwohl diese Tabelle nach den gleichen Grundsätzen aufgestellt wurde, wie Tabelle 3, d. h. trotz der Gleichheit der Q_1 -Werte und der ungefähren Übereinstimmung der Q_0 -Werte, lassen sich wegen der Klassenbildung der Q_0 -Werte und der veränderlichen Größe der Rückweisungsrisiken in Tabelle 3 keine weiteren Schlüsse aus dem Vergleich mit Tabelle 4 ziehen.

²⁾ a. a. O., s. auch *Vance/Neter*, a. a. O., S. 93 f.

Auswirkung der Verringerung von β so gering ist, daß der Stichprobenumfang davon praktisch nicht berührt wird. *Jede Verringerung eines der Risiken ruft grundsätzlich eine Vergrößerung des Stichprobenumfangs hervor.* Erst recht muß die gleichzeitige Einschränkung beider Risiken zu einer Steigerung des Stichprobenumfangs führen.

Nicht eingegangen wurde bisher auf die *Wirkung der Art des Auswahlverfahrens auf den Stichprobenumfang*¹⁾. Die Betrachtung der Tabelle 3 zeigt, daß die kumulierten Stichprobenumfänge der zweiphasigen Annahmepläne erheblich größer sind als die Umfänge der einphasigen Pläne für die gleichen Bedingungen. Es *kann* aber auch schon bei wesentlich geringeren Stichprobenumfängen, nämlich nach der ersten Phase, eine endgültige Entscheidung fallen (was allerdings nicht aus der Tabelle 3 hervorgeht, da die zugehörigen Annahme- und Rückweisungszahlen dort nicht aufgeführt sind). Zwar läßt sich im Einzelfall nicht vorhersagen, wann, d. h. in welcher Auswahlphase, die Entscheidung über Annahme oder Rückweisung der Grundgesamtheit fallen wird, der *Vorteil mehrphasiger Annahmestichproben und Sequenzstichproben* gegenüber den einphasigen Stichprobenverfahren liegt aber darin, daß der *mittlere Stichprobenumfang* — d. i. der Umfang, bei dem im Durchschnitt eine Entscheidung über die Annahme oder die Rückweisung fällt, wenn man denselben Prüfplan sehr oft unter den gleichen Bedingungen anwendet — in der Regel *mit zunehmender Zahl der geplanten Phasen kleiner wird*. Der Wert dieser Verfahren liegt ferner darin, daß der mittlere Stichprobenumfang außerdem vom tatsächlichen (aber unbekannten) Fehleranteil Q der Grundgesamtheit abhängt. Der mittlere Stichprobenumfang ist in der Regel am größten, wenn der wahre Fehleranteil Q größer als der durchschnittliche Fehleranteil Q_0 , aber kleiner als die Rückweisungsgrenze Q_1 ist, d. h. wenn $Q_0 < Q < Q_1$; der mittlere Stichprobenumfang wird im allgemeinen desto kleiner sein, je mehr sich Q von dem kritischen Bereich zwischen Q_0 und Q_1 entfernt.

c) Anwendbarkeit auf das Prüfungsgebiet

(1) Prüferische Auswertung

Als Ergebnis dieser Untersuchungen kann festgestellt werden, daß der Stichprobenumfang desto größer sein muß, je trennschärfer die Prüfplankurve, d. h. je größer Q_0 und Q_1 bei gegebener Differenz ($Q_1 - Q_0$) und gegebenen Risiken α und β , ferner auch je kleiner der Abstand ($Q_1 - Q_0$) bei gegebenen Risiken α und β und je kleiner schließlich α und β bei gegebenen Werten für Q_0 und Q_1 , ist. Ist der durchschnittliche Fehleranteil Q_0 extrem klein und wird auch die Rückweisungsgrenze Q_1 sehr nahe bei Null festgesetzt, so ist die Kleinheit des Abstandes ($Q_1 - Q_0$) für den Stichprobenumfang weit stärker maßgebend als die Kleinheit von Q_0 und Q_1 , so daß der Rationalisierungseffekt der Stichprobe gegenüber der Vollerhebung unter Umständen minimal wird. Schließlich müssen dann entweder von vornherein Vollerhebungen stattfinden oder die in der Grundgesamtheit enthaltenen Fehler mit anderen Mitteln als Stichproben aufgedeckt werden. Als solche kommen dann insbesondere die laufenden Kontrollen im Rahmen des innerbetrieblichen Kontrollsystems in Betracht.

¹⁾ Zu diesem Absatz s. Vance/Neter, a. a. O., S. 48 ff., 108 f., 115.

Der für eine abschließende Entscheidung erforderliche Stichprobenumfang kann in der Regel durch Planung einer größeren Anzahl von Auswahlphasen oder einer Sequenzstichprobe vermindert werden und wird erst recht dann sehr klein sein, wenn die tatsächliche Qualität der Grundgesamtheit wesentlich von den dem Prüfplan zugrunde gelegten Fehleranteilen Q_0 und Q_1 abweicht.

In der Prüfungspraxis kommen *Verstöße gegen die gesetzlichen Ordnungsvorschriften in unterschiedlichem Umfang* vor, z. B. wird das Erfordernis der lebenden Sprache oder der Wertangaben in der Landeswährung praktisch immer eingehalten. Dagegen lassen sich gelegentlich Verstöße z. B. gegen das Belegprinzip oder das Verbot von Rasuren oder von Bleistifteintragungen beobachten. Aber diese Verstöße müssen schon einen gewissen Umfang erreichen, um den Prüfer zum Abbruch der Prüfung oder zur Einschränkung oder Verweigerung des Bestätigungsvermerks zu veranlassen. Sie führen dagegen in der Regel zu einer Intensivierung der Prüfung der materiellen Richtigkeit der Buchführung.

Da hier also die Bedingungen für die Bestimmungsgrößen der Annahmestichproben keine besondere Schärfe anzunehmen brauchen, wird man, wie aus den obigen Betrachtungen hervorgeht, *in der Regel bereits mit verhältnismäßig kleinen Stichprobenumfängen hinreichende Sicherheit erhalten können, ob die Buchführung in ausreichendem Maße den gesetzlichen Ordnungsvorschriften genügt oder nicht.*

(2) Grundlagen für die Festsetzung der Bestimmungsfaktoren

Umfang und durchschnittlicher Fehleranteil der Grundgesamtheit

Bei den Annahmeprüfungen findet sich die *Grundgesamtheit* des Prüfungsstoffs in den Büchern und Schriften der Gesellschaft, den Belegen, Listen usw., soweit ihre Führung gesetzlich vorgeschrieben ist. In erster Linie beziehen sich diese Vorschriften auf die *Geschäftsbuchführung* und nur insofern auf die Betriebsbuchführung, als deren Ergebnisse für die ordnungsmäßige Geschäftsbuchführung benötigt werden. Innerhalb der gesetzlich vorgeschriebenen Geschäftsbuchführung ist es aber grundsätzlich gleichgültig, wo Verstöße gegen die gesetzlichen Ordnungsvorschriften auftreten, ob z. B. Rasuren im Hauptbuch, auf den Konten oder Unterkonten oder in den Journalen vorkommen. Insoweit kann daher die gesetzlich vorgeschriebene Geschäftsbuchführung als *eine* Grundgesamtheit angesehen werden. Umgekehrt ist es grundsätzlich notwendig, daß jede Untersuchungseinheit allen gesetzlichen Ordnungsvorschriften entspricht.

Die *Untersuchungseinheit* ist der Geschäftsvorfall bzw. die Dokumentation des Geschäftsvorfalles, z. B. in Form eines Belegs oder der Eintragung (Buchung) im Konto oder Journal. Die *Untersuchungsmerkmale* sind ganz allgemein die Einhaltung und die Verletzung der jeweiligen Ordnungserfordernisse.

Bereits die obigen Darstellungen lassen erkennen, daß *kleine* Grundgesamtheiten relativ große Stichprobenumfänge erfordern, daß der absolute Stichprobenumfang dagegen bei *großen* Grundgesamtheiten nur unwesentlich und bei *sehr großen* Grundgesamtheiten überhaupt nicht von der genauen Größe der Grundgesamtheit abhängt. Für die Planung von Annahmestichproben genügt daher bei größeren Grundgesamtheiten eine *überschlägige Abschätzung ihres Umfangs*.

Schwieriger als die Definition der Grundgesamtheit, die richtige Abschätzung des Umfangs dieser Grundgesamtheit, die Definition der Untersuchungseinheiten und ihrer Untersuchungsmerkmale, ist für die Planung angemessener Stichprobenverfahren die *richtige Feststellung des durchschnittlichen Fehleranteils* Q_0 . Ihr kommt größere Bedeutung zu, weil durch die Festsetzung von Q_0 und die im übernächsten Abschnitt zu behandelnde Festsetzung der Rückweisungsgrenze Q_1 auch die absolute und relative Größe des kritischen Bereichs bestimmt wird. Der Stichprobenumfang muß aber, wie gezeigt, um so größer sein, je kleiner der Bereich zwischen Q_0 und Q_1 und je größer der durchschnittliche Fehleranteil Q_0 sind. Bei der Planung des Stichprobenverfahrens sollte daher dieser Fehleranteil möglichst wirklichkeitsnah angesetzt werden.

Liegen *Anhaltspunkte für die Größe des Fehleranteils* vor, z. B. aus Informationen des Unternehmens, insbesondere der Innenrevision oder aufgrund der vom Prüfer im Vorjahr ermittelten Prüfungsergebnisse, so kann man sich bei der Planung der Stichprobe auf diese Auskünfte stützen. Dabei ergeben sich keine Schwierigkeiten, wenn alle Informationen gleichlautend sind.

Liegen *ungenau oder wenig unterschiedliche Informationen* über die zu vermutende Größe des Fehleranteils vor, so ist derjenige Wert als Ausgangsbasis der Planungsberechnungen zu benutzen, der der Rückweisungsgrenze Q_1 am nächsten kommt. Unter diesen Umständen ist eine Fehlentscheidung am unwahrscheinlichsten, weil dann ceteris paribus die Kriterien des Stichprobenplans am schärfsten sind.

Liegen *keine Informationen* vor und sind auch keine befriedigenden Auskünfte zu erlangen, so kann der Fehleranteil durch Vorstichproben überschlägig abgeschätzt werden. Diese Vorstichproben können gleichzeitig als erste Phasen der Hauptstichprobe gelten.

Auswahlgrundlagen und Auswahlverfahren

Als *Auswahlgrundlagen* kommen die Belegablagen, Journale, Konten, Kontenverzeichnisse usw. in Frage. Welche davon im Einzelfall praktische Bedeutung erlangen, richtet sich vornehmlich nach der Buchführungsorganisation und der vom Prüfer gewünschten Prüfungsrichtung (progressiv — retrograd). Z. B. können für die Prüfung der Rasuren das Kontenverzeichnis mit der Angabe sämtlicher Konten und der Zahl der innerhalb der Konten benutzten Kontenblätter und bei der Prüfung der Belege eine Liste der Belegordner mit Angabe der in ihnen enthaltenen Zahl der Belege die Auswahlgrundlagen sein.

Unter den *Auswahlverfahren* sind die *mehrphasigen* Verfahren und die Verfahren der *Sequenzstichproben* aus den auf Seite 34 dargelegten Gründen bevorzugt anzuwenden. Über die Zahl der zu planenden Phasen entscheiden — soweit es sich nicht um Sequenzstichproben handelt — praktische Gesichtspunkte¹⁾. Im Rahmen der Abschlußprüfung sind z. B. die Kostenunterschiede zwischen verschiedenen Stichprobenplänen zu beachten, die den gleichen prüferischen Bedingungen genügen, sich aber hinsichtlich der geplanten Phasenzahl bzw. des mittleren Stichprobenumfangs unterscheiden. Ferner ist das prüferische Interesse, den tatsächlichen Fehleranteil zu kennen, von Einfluß, weil die Qualität der Buchführung als Maßstab für den Sicherheitsgrad bei den Planungen der materiellen

¹⁾ Vance/Neter, a. a. O., S. 50 ff.

Stichprobenprüfungen zugrunde gelegt werden kann. Deshalb sollte die *erste Phase keinesfalls zu klein angesetzt werden*. Vance/Neter geben bei dieser Zielsetzung einem zweiphasigen Auswahlverfahren den Vorzug vor den ein- oder den mehr als zweiphasigen Verfahren für gleiche prüferische Bedingungen¹⁾.

Nicht zuletzt ist es bei dem hier behandelten Erhebungsgegenstand auch psychologisch in der Regel befriedigender, daß die endgültige Entscheidung erst in einer fortgeschrittenen Phase der Stichprobenprüfung gefällt wird, während man eine Entscheidung aufgrund einer einphasigen Prüfung leicht als „voreilig getroffen“ empfindet.

Die *Bildung und Abgrenzung von Auswahlgesamtheiten*, d. h. eine Untergliederung der Grundgesamtheit in mehrere Teilgesamtheiten zu Auswahlzwecken, ist bei der Anwendung von Annahmestichproben aus den auf Seite 35 dargelegten Gründen nicht sinnvoll. Eine Aufteilung der Annahme- und Rückweisungszahlen auf die einzelnen Teilgesamtheiten führt nicht zu dem prüferischen Ziel, eine klare Entscheidung über Annahme oder Rückweisung der ganzen Grundgesamtheit zu erlangen.

In den *einzelnen Auswahlphasen* sind daher die Stichproben des jeweils vorgeschriebenen Umfangs durch *uneingeschränkte Zufallsauswahl* zu entnehmen.

Risiken und Rückweisungsgrenze

Die seitens des Prüfers vorzunehmende Festsetzung des *Rückweisungsrisikos α* und des *Annahmerisikos β* ist an das prüferische Ermessen gebunden, solange für diese Risiken noch keine Berufsnormen gebildet worden sind.

Auch bei der Festsetzung der Rückweisungsgrenze Q_1 ist der Prüfer bisher auf seinen Ermessensentscheid angewiesen. Bei ihrer Bestimmung sollte der Prüfer sich jedoch über die Folgen im klaren sein, die er mit einer etwaigen Rückweisung der Grundgesamtheit aufgrund des Stichprobenergebnisses verbinden will (Verweigerung bzw. Einschränkung des Bestätigungsvermerks, Erwähnung im Prüfungsbericht, Intensivierung der Prüfung etc.).

In den *Beispielen* der Literatur liegen die je nach Lage der Dinge individuell für erforderlich gehaltenen Werte für das Rückweisungsrisiko α bei 5%, für das Annahmerisiko β zwischen 5% und 10% und für die Rückweisungsgrenze Q_1 zwischen 3% und 9%²⁾.

Die *gemeinsame Prüfung mehrerer Untersuchungsmerkmale* im gleichen Annahmestichprobenverfahren, wie es bei der Prüfung der gesetzlichen Ordnungsvorschriften der Fall ist, wirft die Frage auf, wie Unterschiede in der Art, der Bedeutung und der Häufigkeit der Fehler in den Stichprobenplänen berücksichtigt werden können. Eine praktikable Lösung dieses Problems sieht Mackenzie³⁾ darin, „die Fehlertypen nach ihrer Bedeutung zu klassifizieren, für jede Klasse annehmbare und unannehmbare Qualitäten (d. s. der durchschnittliche Fehleranteil Q_0 und die Rückweisungsgrenze Q_1 , Anm. d. Verf.) festzusetzen und diese Qualitäten zu einer Art Hindernisbahn zu machen, die eine Gruppe von Dokumenten bewältigen muß, um angenommen zu werden. In gewissem Sinn bedeutet das,

¹⁾ a. a. O., S. 52

²⁾ Vgl. Vance/Neter, a. a. O., S. 134, 136.

³⁾ Mackenzie, H. C., Statistical Sampling in Accounting; The Accountant, 1. 11. 1958, S. 524.

jede Fehlerart so zu behandeln, als hätte sie ein anderes Maß¹⁾). Das Dokument der Rechnungslegung würde dann zu Prüfungszwecken wie ein Industrieprodukt behandelt, das verschiedene Eigenschaften haben muß, ehe es als befriedigend erachtet werden kann“.

In ähnlichem Sinn äußern sich auch *Vance/Neter*. Sie stellen gleichfalls fest, daß es allgemeine Übung sei, eine Grundgesamtheit nur dann anzunehmen, wenn aufgrund der Prüfpläne die Grundgesamtheit hinsichtlich aller Fehlertypen angenommen werden kann²⁾).

Die Frage, wie diese *Fehlertypen bezüglich ihres relativen Gewichts einzustufen* sind, hat bisher keine befriedigende Antwort gefunden³⁾.

Solange noch kein objektiver Maßstab für die Bedeutung der Fehlerarten (z. B. Rasur, Bleistifteintragung, fehlender Beleg, falsche Numerierung) für die Sicherheit der Rechnungslegung gefunden ist, muß es dem *prüferischen Ermessen* überlassen bleiben, in welchem Umfang Fehler der verschiedenen Typen in der Grundgesamtheit zugelassen werden können, m. a. W. welche Rückweisungsgrenzen für die einzelnen Fehlertypen vorzuschreiben sind.

Demgegenüber ist der *Wert bzw. Geldbetrag des Fehlers* bei der Prüfung der formellen Ordnungserfordernisse nicht von entscheidender Bedeutung. Eine Rasur in der Spalte für DM 1 000, — hat dieselbe Bedeutung wie eine solche in der Pfennigspalte.

Wenn auch die Annahmestichproben bei der Abschlußprüfung nur für die Prüfung der Einhaltung der gesetzlichen Ordnungsvorschriften von unmittelbarer Bedeutung sind, so soll hier doch hervorgehoben werden, daß sie große Bedeutung für die *Innenrevision* besitzen. Hier geht es um die laufende Überwachung der Arbeitsqualität in der Führung der Bücher und der Einhaltung der durch das innerbetriebliche Kontrollsystem vorgegebenen Arbeitsabläufe, im Falle der Überschreitung einer Rückweisungszahl, d. h. bei Rückweisung einer Grundgesamtheit, um die Ermittlung der Fehlerquellen⁴⁾. Diese Überwachung kann durch Annahmestichproben auf sehr rationelle Weise erfolgen, wie zahlreiche Beispiele in der Literatur⁵⁾ beweisen.

B. Prüfung der Einhaltung der unternehmensinternen Ordnungsvorschriften

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Die unternehmensinternen Ordnungsvorschriften haben den Zweck, das Unternehmensvermögen zu schützen, die Zuverlässigkeit der Buchführung zu gewährleisten und schließlich die Einhaltung der Unternehmenspolitik zu fördern⁶⁾. Sie gehen damit weit über die vom Gesetzgeber erhobenen Ordnungsforderungen hinaus. Andererseits schließen sie sie aber auch in sich ein, soweit sie der Sicherheit der Rechnungslegung dienen.

1) Hervorhebung vom Zitierenden.

2) *Vance/Neter*, a. a. O., S. 75.

3) Zu diesem und dem folgenden Absatz s. auch *Vance/Neter*, a. a. O., S. 126 ff.

4) Vgl. *Mackenzie*, a. a. O., S. 525.

5) Vgl. *Vance/Neter*, a. a. O.; *Trueblood/Cyert*, a. a. O., u. v. a.

6) Definition des amerikanischen Instituts der Wirtschaftsprüfer; s. *Pougin*, a. a. O., S. 15.

Die Prüfungen sollen ergeben, ob und in welchem Umfang diese Vorschriften beachtet bzw. verletzt worden sind, z. B. ob und in welchem Umfang für die einzelnen Buchungen Belege vorhanden, die Belege von den jeweils zuständigen Personen ausgestellt und unterschrieben, die vorgeschriebenen Kontrollen durchgeführt worden sind usw.

Während es sich bei den gesetzlichen Ordnungsvorschriften immer nur um Einzelgebote oder Einzelverbote handelt, die zwar letztlich alle demselben Ziel dienen, können die unternehmensinternen Ordnungsvorschriften bis zu einem mehr oder minder geschlossenen System entwickelt sein. Hier ist dann der Grad der Sicherheit der Rechnungslegung sowohl von der Güte und Zweckmäßigkeit des Systems als auch vom Umfang der tatsächlichen Anwendung der verschiedenen Kontrollmaßnahmen abhängig.

Die Prüfung der Einhaltung des innerbetrieblichen Kontrollsystems umfaßt also zwei Fragen

- a) Welchen Umfang hat das innerbetriebliche Kontrollsystem und wie ist seine Zweckmäßigkeit zu beurteilen?
- b) Wie groß ist die tatsächliche Wirksamkeit des innerbetrieblichen Kontrollsystems?

Die *erste Frage* nach der Qualität des Systems kann nur mittels systematischer Untersuchungen, wie sie z. B. *Pougin*¹⁾ beschreibt, nicht aber mit Hilfe von Stichprobenprüfungen beantwortet werden. *Pougin* unterscheidet Untersuchungen mit Hilfe des Internal-Control-Fragebogens, mit Hilfe von Arbeitsablauf-Schaubildern und andere Untersuchungstechniken, wie z. B. Prüfungen an Hand von Merklisten, Formularsammlungen, Organisationsplänen oder beschreibenden Darstellungen des innerbetrieblichen Kontrollsystems usw.

Die *zweite Frage* läßt sich dagegen nicht durch Untersuchungen der vorgenannten Art beantworten.

Die Erledigung der Geschäftsvorfälle wird in einer Kette aufeinanderfolgender und aufeinanderaufbauender Geschäftshandlungen vollzogen. Das innerbetriebliche Kontrollsystem schreibt die Reihenfolge der Handlungen vor und soll dafür sorgen, daß diese Reihenfolge eingehalten wird. Will der Prüfer die Wirksamkeit des Systems feststellen, so kann er die einzelnen Geschäftshandlungen darauf untersuchen, *in welchem Umfang* bei ihnen *Verstöße gegen das innerbetriebliche Kontrollsystem* zu beobachten sind, d. h., aus der relativen Häufigkeit des Auftretens fehlerhafter Einheiten kann auf die Wirksamkeit des innerbetrieblichen Kontrollsystems geschlossen werden.

Hierbei ist zunächst unterstellt, daß alle Verstöße gegen die Ordnungsvorschriften für die Gewährleistung der Sicherheit der Rechnungslegung gleichwertig sind. Tatsächlich sind *Fehlerart* und *Fehlerort* jedoch von unterschiedlicher Bedeutung. Ein Verstoß gegen das Prinzip der Funktionsteilung, also z. B. die Anweisung und Auszahlung eines Barbetrages durch dieselbe Person, ist von größerer Bedeutung als z. B. das Fehlen einer Telefonrechnung, die mangelhafte Kontrolle bei der Warenannahme im Lager wichtiger als die mangelnde Kontrolle bei der Annahme von Büromaterialien.

¹⁾ a. a. O., S. 83 ff.

Die Wirksamkeit des gesamten Kontrollsystems kann man also nicht ohne weiteres an der Zahl der insgesamt vorgekommenen Verstöße beurteilen. Mit Rücksicht auf die unterschiedliche Bedeutung von Fehlerart und Fehlerort muß man sich vielmehr darauf beschränken, die Wirksamkeit der einzelnen Kontrollvorschriften zu prüfen. In diesem Fall ist dann sichergestellt, daß Fehlerart und Fehlerort als gleichwertig anzusehen sind, so daß sich die Wirksamkeit einer solchen Einzelvorschrift quantitativ an der Zahl der vorgekommenen Verstöße messen läßt.

Über den *Umfang der Stichproben* zur Feststellung der tatsächlichen Wirksamkeit des innerbetrieblichen Kontrollsystems äußert sich *Pougin*¹⁾ wie folgt:

„Die Untersuchung des Kontrollsystems in großen Unternehmen ist eine umfangreiche Aufgabe. Der Prüfer muß sich auch hier wie bei der eigentlichen Abschlußprüfung auf Stichproben beschränken. Das wirft die Frage auf: Wie groß ist der durchschnittliche Umfang der Stichproben, der für die Feststellung der tatsächlichen Wirksamkeit des internen Kontrollsystems bei einer normalen Abschlußprüfung erforderlich ist?

Jeder Versuch, diesen Umfang in Prozenten des Umfangs der Stichproben, die für die eigentliche Abschlußprüfung erforderlich sind, auszudrücken, führt zu einem unbefriedigenden Ergebnis, wenn man auch sagen kann — wie ein Versuch des amerikanischen Instituts der Wirtschaftsprüfer zeigt —, daß es wahrscheinlich nicht weniger als 10% und nicht mehr als 25% sind.

Der Prüfer sieht sich vor dieselbe Aufgabe gestellt, wie bei der Bestimmung des Prüfungsumfangs der eigentlichen Abschlußprüfung. Die Lösung kann ihm nur seine berufliche Urteilskraft in Übereinstimmung mit den bestehenden Berufsgrundsätzen geben. Was bleibt, ist die Verpflichtung festzustellen, in welchem Maß er sich bei der Planung seiner Prüfung auf das bestehende interne Kontrollsystem verlassen kann.“

Während *Pougin* also die Bestimmung des Stichprobenumfangs praktisch ausschließlich in das prüferische Ermessen stellt, kann die Stichprobentheorie die Anteile der Einheiten mit bestimmten Eigenschaften (fehlerfrei — fehlerhaft) unter bestimmten Voraussetzungen mit gewünschter Sicherheit und Genauigkeit auf rationelle Weise objektiv ermitteln. Es soll nunmehr geprüft werden, welche Voraussetzungen gegeben sein müssen, um solche *Schätzstichproben* zur Prüfung heranzuziehen.

b) Schätzung von Fehleranteilen mittels Schätzstichproben

(1) Formelapparat

Bezeichnet man den Mittelwert aller möglichen Schätzwerte q für den Fehleranteil Q mit $\bar{q}$, so ist²⁾

$$(17) \quad \bar{q} = Q$$

Ferner besteht eine Beziehung zwischen der *Streuung der Schätzwerte* um diesen Mittelwert und der Streuung der Merkmalwerte der Untersuchungseinheiten der Grundgesamtheit. Als Maß für diese Streuungen wird die „*Varianz*“ gewählt,

1) a. a. O., S. 92.

2) S. z. B. *Kellerer* (1953), S. 36.

weil sie, wie sich später zeigen wird, als Maßstab für die Größe des Zufallsfehlers besonders geeignet ist.

Die Varianz einer Wertereihe ist der Durchschnitt der Quadrate der Differenzen zwischen den Merkmalwerten dieser Reihe und ihrem Mittelwert. Diese Definition trifft auch auf Reihen zu, deren Einheiten sich durch bestimmte Eigenschaften auszeichnen, da man den jeweils alternativen Eigenschaften (z. B. Fehlerfreiheit — Fehlerhaftigkeit) die Merkmalwerte Null resp. Eins geben kann. Bezeichnet man den Umfang der Grundgesamtheit mit N , die Zahl der darin enthaltenen fehlerhaften Einheiten mit M und ihren Anteil mit Q und den Anteil der fehlerfreien Einheiten in der Grundgesamtheit mit P , so sind, wenn μ das arithmetische Mittel darstellt

$$\mu = \frac{1 M + 0 (N - M)}{N} = \frac{M}{N} = Q$$

und wenn σ^2 die Varianz bezeichnet,

$$\sigma^2 = \frac{1^2 M + 0^2 (N - M)}{N} - \mu^2 = \frac{M}{N} - Q^2 = Q (1 - Q) = P Q$$

Bezeichnet man die Varianz aller möglichen q -Werte mit σ_q^2 , so sind bei uneingeschränkter Zufallsauswahl²⁾,

wenn N sehr groß gegenüber n ,

$$(18) \quad \sigma_q^2 = \frac{P Q}{n}$$

wenn N nicht sehr groß gegenüber n ,

$$(19) \quad \sigma_q^2 = \frac{P Q}{n} \frac{N - n}{N - 1}$$

Im zweiten Fall tritt also zu dem vorgenannten Verhältnis noch der „*Endlichkeitsfaktor*“ $\frac{N - n}{N - 1}$ hinzu. Dieser mißt praktisch den Anteil der nicht in die Stichprobe gezogenen Einheiten am Umfang der Grundgesamtheit.

Bei Stichprobenerhebungen sind die Anteilwerte P und Q der Grundgesamtheit in der Regel nicht bekannt. Als *Näherungswerte* werden für den Fehleranteil Q der Fehleranteil q der Stichprobe, entsprechend p für P und für die Varianz $P Q$ die Varianz $p q$ der Stichprobe genommen.

Bei *komplizierteren Auswahlverfahren* (Schichtung, Stufung etc.) wird die Varianz der möglichen q -Werte außer durch die Struktur der Grundgesamtheit auch durch die Besonderheiten der Auswahl bestimmt.

Diese Formeln haben daher den der Kompliziertheit des Auswahlverfahrens entsprechenden komplizierten Aufbau.

¹⁾ S. z. B. Kellerer (1953), S. 36.

²⁾ S. z. B. Kellerer (1953), S. 36.

Werden z. B. in einem *zweistufigen Auswahlverfahren* auf der ersten Auswahlstufe von N' Primäreinheiten n' ausgewählt und umfaßt die i . Primäreinheit N_i Sekundäreinheiten, von denen n_i in die Stichprobe gelangen¹⁾, so ist, wenn in der i . Primäreinheit der Anteil fehlerhafter Sekundäreinheiten P_i und der Anteil fehlerfreier Sekundäreinheiten Q_i betragen, die Varianz σ_q^2 der Fehleranteile q in den möglichen Stichproben gegeben durch²⁾:

$$(20) \quad \sigma_q^2 = \frac{1}{N^2} \frac{N'}{n'} \left[\frac{N' - n'}{N' - 1} \sum_{i=1}^{N'} \left(N_i Q_i - \frac{N}{N'} Q \right)^2 + \sum_{i=1}^{N'} \frac{N_i^2 P_i Q_i}{n_i} \frac{N_i - n_i}{N_i - 1} \right]$$

Hierin sind als Grenzfälle die Varianz der Schätzwerte beim geschichteten Stichprobenverfahren und beim Klumpenauswahlverfahren enthalten. Beim geschichteten Verfahren werden die Primäreinheiten voll erhoben ($n' = N'$). Dann fällt der erste Summand in der Klammer weg und $N'/n' = 1$. Im Fall der Klumpenauswahl beträgt der Auswahlatz unter den Sekundäreinheiten 100%, d. h., es ist jeweils $n_i = N_i$, wobei der zweite Summand in der Klammer fortfällt.

Als *Näherungswerte* für die unbekannten P_i und Q_i werden — wie bei der uneingeschränkten Auswahl — die jeweiligen Stichprobenanteile p_i und q_i genommen.

Es ist nun eine Beziehung zwischen der Varianz aller gleichmöglichen Schätzwerte und der *Häufigkeitsverteilung* dieser Schätzwerte herzustellen, damit die Häufigkeiten von Abweichungen bestimmter Größe zwischen den Schätzwerten und dem gesuchten Wert berechnet werden können.

Nach dem „zentralen Grenzwertsatz der Wahrscheinlichkeiten“ nähert sich die Häufigkeitsverteilung der arithmetischen Mittelwerte — und damit verwandter Werte, wie Gesamtwerte oder Anteilwerte — aller Stichproben gleichen Umfangs aus ein und derselben Grundgesamtheit um so stärker einer *Gaußschen Normalverteilung*, je größer der Umfang der Stichprobe ist³⁾.

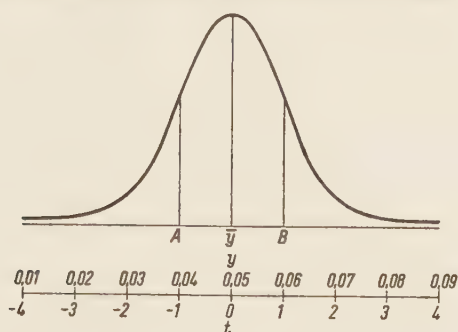
Eine *Gaußsche* Normalverteilung hat glockenförmige, symmetrische Gestalt (vgl. Figur 2). Sie ist durch den Mittelwert der Merkmalwerte, hier der Schätzwerte, und ihre mittlere quadratische Abweichung — sie ist die Quadratwurzel aus der Varianz und heißt auch Standardabweichung — eindeutig festgelegt. Der von den in zwei Punkten (z. B. A und B in Figur 2) auf der Abszisse errichteten Senkrechten, der Abszisse und der Kurve umschlossene Anteil der Gesamtfläche zwischen Abszisse und Kurve ist ein Maß für die relative Häufigkeit (Anteil) der Stichproben, deren Schätzwert in dem durch diese Punkte abgegrenzten Wertebereich liegt.

1) $\sum_{i=1}^{N'} N_i = N$; $\sum_{i=1}^{n'} n_i = n$

2) Vgl. Kellerer (1953), S. 148.

3) Z. B. Anderson, O., Probleme der statistischen Methodenlehre in den Sozialwissenschaften, 3. Aufl., Würzburg 1957, S. 123, 136 f.; auch der nicht ausgewählte Rest ($N - n$) der Grundgesamtheit muß genügend groß sein, damit die Zahl $\binom{N}{n}$ der möglichen Stichproben gleichen Umfangs nicht zu klein wird; s. Kellerer (1953), S. 27.

Sind diese Punkte (*A* und *B*) so bestimmt, daß sie, wie in Figur 2, mit gleichem Abstand beiderseits des Mittelwertes aller Schätzwerte liegen, und wird dieser ihr



Figur 2: Gaußsche Normalverteilung ($\bar{y} = 0,05$; $\sigma_y = 0,01$)

Abstand als ein Vielfaches der Standardabweichung ausgedrückt, so hängt der Flächenanteil nur noch von der Größe dieses Vielfachen ab¹⁾. Tabelle 5 zeigt einige Werte dieser Funktion, wobei t das Vielfache und $F(t)$ den Flächenanteil bezeichnen.

Tabelle 5: Flächenanteil $F(t)$, der von den im Abstand des t -fachen der Standardabweichung vom Mittelwert der Normalverteilung auf der Abszisse errichteten Senkrechten, der Kurve und der Abszisse umschlossen wird, an der Gesamtfläche zwischen Abszisse und Kurve bei bestimmten Werten von t .

t	0,67	1,00	1,28	1,44	1,64	1,96	2,00	2,58	3,00	4,00
$F(t)$	0,500	0,683	0,800	0,850	0,900	0,950	0,955	0,990	0,997	0,99 994

Bezeichnet man den Mittelwert aller gleichmöglichen Schätzwerte mit $\bar{y}$ und ihre Standardabweichung mit σ_y , so liegen bei nicht zu kleinem Stichprobenumfang und genügender Größe der Grundgesamtheit z. B. 95,5% aller möglichen Stichprobenergebnisse zwischen $\bar{y} - 2\sigma_y$ und $\bar{y} + 2\sigma_y$.

Ebenso kann die Beziehung zwischen $F(t)$ und t dazu benutzt werden, die Wahrscheinlichkeit für die Auswahl einer Stichprobe anzugeben, deren Ergebnis (Schätzwert) innerhalb des durch ein bestimmtes Vielfaches der Standardabweichung σ_y beiderseits des Mittelwertes $\bar{y}$ abgegrenzten Bereichs liegt, also z. B. daß mit 99,7% Wahrscheinlichkeit das Stichprobenergebnis im Bereich zwischen $\bar{y} - 3\sigma_y$ und $\bar{y} + 3\sigma_y$ liegen wird. Diese Wahrscheinlichkeit $F(t)$ bezeichnet man als *Sicherheitsgrad*, während für den Wertebereich die Bezeichnung *Vertrauensbereich* üblich ist.

Handelte es sich im letzten Beispiel um den „direkten Schluß“ von den Werten der Grundgesamtheit auf die der Stichprobe — auch Inklusionsschluß genannt —, so läßt sich aufgrund mathematischer Überlegungen²⁾ auch der „Repräsentations-schluß“ von den Werten einer Stichprobe auf die der Grundgesamtheit ziehen:

1) $F(t) = \frac{1}{2\pi} \int_{-t}^{+t} e^{-\frac{1}{2}u^2} du$, wobei $u = \frac{y - \bar{y}}{\sigma_y}$; $\pi = 3,14159 \dots$; $e = 2,71828 \dots$

vgl. Anderson, a. a. O., S. 111.

2) Anderson, a. a. O., S. 115 f.

Bezeichnet man die zulässige Größe der Differenz $|y - \bar{y}|$ mit e (Größe des Vertrauensbereichs $= 2e$), so ist zunächst das Vielfache der Standardabweichung σ_y definiert durch

$$(21) \quad t = \frac{|e|}{\sigma_y}$$

und bei nicht zu kleinem Stichprobenumfang und genügender Größe der Grundgesamtheit erhält man mit einem durch $F(t)$ gegebenen Sicherheitsgrad einen Stichprobenwert y , für den die Beziehung

$$(22) \quad y - ts_y \leq \bar{y} \leq y + ts_y$$

zutritt, wobei s_y der Schätzwert für σ_y ist.

Sind für den Sicherheitsgrad und den Vertrauensbereich und damit auch für e und t bestimmte Werte vorgeschrieben, so können diese Bedingungen offenbar nur eingehalten werden, wenn

$$(23) \quad \sigma_y \leq \frac{|e|}{t}$$

Nun ist aber im Fall der Anteilsschätzung $\sigma_y = \sigma_q$ und die Varianz der Schätzwerte q bei uneingeschränkter Zufallsauswahl durch die Formeln (18) bzw. (19) gegeben. Nach Einsetzen von (18) bzw. (19) in (23) und Auflösung der Formeln nach n erhält man die Planungsformeln für den Stichprobenumfang.

Diese lauten
bei Vernachlässigung des Endlichkeitsfaktors

$$(24) \quad n \geq \frac{t^2 P Q}{e^2}$$

bei Berücksichtigung des Endlichkeitsfaktors

$$(25) \quad n \geq \frac{t^2 P Q N}{e^2 (N - 1) + t^2 P Q}$$

(2) Wirkung der Bestimmungsfaktoren auf den Stichprobenumfang

Es ist nun zu fragen, wie sich

- a) der Umfang der Grundgesamtheit,
- b) die Größe der Varianz für das Untersuchungsmerkmal,
- c) die Art des Auswahlverfahrens,
- d) die Größe von Sicherheitsgrad und Vertrauensbereich

ceteris paribus auf den erforderlichen *Mindestumfang der Stichprobe* auswirken.

Der *Umfang der Grundgesamtheit* spielt, vergleicht man die beiden obigen Formeln (24) und (25), nur bei Berücksichtigung des Endlichkeitsfaktors eine Rolle. Der Einfluß des Umfangs der Grundgesamtheit auf den Stichprobenumfang läßt sich beurteilen, wenn man das Verhältnis zwischen dem gemäß Formel (24) erforderlichen Stichprobenumfang — er sei hier mit n_1 bezeichnet — und dem unter den gleichen Bedingungen gemäß Formel (25) erforderlichen Umfang der Stichprobe — n_2 — vergleicht.

Man erweitert zu diesem Zweck Zähler und Nenner von (25) mit e^2 und erhält nach Umformung und Einsetzen von (24) die Relation

$$(26) \qquad n_2 = n_1 \frac{N}{(N-1) + n_1} \approx \frac{n_1}{1 + \frac{n_1}{N}} = \frac{n_1}{1 + f}$$

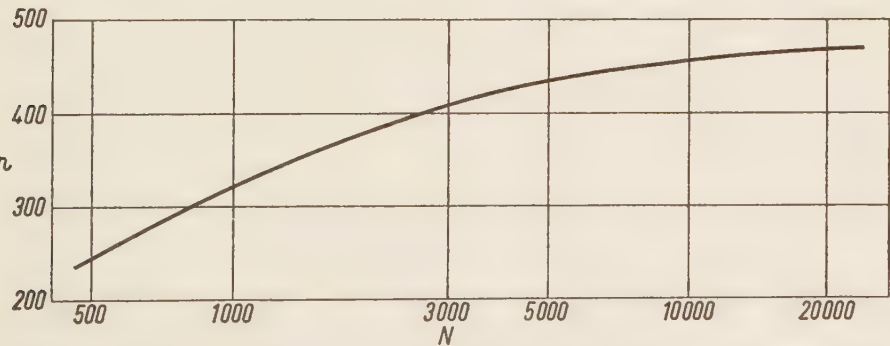
wobei f der Auswahlatz gemäß Formel (24) ist.

Beträgt also z. B. der Anteil der fehlerhaften Einheiten in einer Stichprobe 5% und ist folglich die Varianz $p\,q = 0,95 \cdot 0,05 = 0,0475$, so muß bei $t = 2$ und $e = 0,02$ die Stichprobe gemäß (24) mindestens 475 Einheiten umfassen. Die Berücksichtigung des Endlichkeitsfaktors führt dagegen gemäß sowohl (25) wie (26) zu den in Tabelle 6 angegebenen Werten.

Tabelle 6: Erforderlicher Stichprobenumfang — absolut und relativ — bei Berücksichtigung des Umfangs der Grundgesamtheit bei einem Fehleranteil von $Q = 0,05$ und einem Sicherheitsgrad, der $t = 2$ entspricht, sowie einem Vertrauensbereich von $e = \pm 0,02$.

Umfang der Grundgesamtheit N	500	1 000	3 000	5 000	10 000	20 000
Stichprobenumfang absolut	244	323	411	427	454	465
in % von N	48,8	32,3	13,7	8,5	4,5	2,33

Diese Tabelle läßt erkennen, daß mit steigendem Umfang der Grundgesamtheit der absolute Mindestumfang der Stichprobe anfänglich rasch, mit weiter wachsenden Grundgesamtheiten langsamer zunimmt, bis er eine *obere Grenze* erreicht (hier: $n = 475$), die er *auch bei noch so großen Grundgesamtheiten nicht mehr überschreitet*. Stellt man diese Entwicklung graphisch dar, so erhält man das in Figur 3 gezeigte Bild.



Figur 3: Erforderlicher absoluter Stichprobenumfang bei wachsendem Umfang der Grundgesamtheit bei $Q = 0,05$; $e = 0,02$; $t = 2$ (Abszisse logarithmisch unterteilt)

Aus der Feststellung, daß der absolute Mindestumfang der Stichproben mit wachsendem Umfang der Grundgesamtheiten nur unterproportional zunimmt und von einer gewissen oberen Grenze ab auch bei noch so großen Grundgesamt-

heiten nicht mehr steigt, folgt, daß eine Stichprobe desto rentabler ist, je größer die Grundgesamtheit ist; umgekehrt folgt daraus, daß die Genauigkeit der Stichprobenergebnisse immer weniger vom Umfang der Grundgesamtheit abhängt, je geringer der Auswahlatz ist.

Die Frage, von welchem Umfang der Grundgesamtheit ab die Berücksichtigung des Endlichkeitsfaktors geboten erscheint, kann beantwortet werden, wenn man die gemäß Formel (25) erforderlichen Stichprobenumfänge dem gemäß Formel (24) benötigten Stichprobenumfang gegenüberstellt. Es ergeben sich dann für das obige Beispiel die aus Tabelle 7 ersichtlichen Unterschiede der Auswahlätze.

Tabelle 7: Die Auswahlätze bei Vernachlässigung und bei Berücksichtigung des Endlichkeitsfaktors für $Q = 0,05$; $e = 0,02$; $t = 2$ und verschiedene N .

Umfang N der Grundgesamtheit	500	1 000	3 000	5 000	10 000	20 000
Auswahlatz in %						
a) bei Vernachlässigung des Faktors $n = 475 = \text{const}$	95,0	47,5	15,8	9,5	4,8	2,38
b) bei Berücksichtigung des Faktors vgl. Tabelle 6	48,8	32,3	13,7	8,5	4,5	2,33

Der Auswahlatz ist bei Vernachlässigung der Faktors stets größer als bei seiner Berücksichtigung. Er sinkt in beiden Fällen bei steigendem Umfang der Grundgesamtheit schnell ab, in der vorstehenden Tabelle im Fall a) von 95,0% auf 2,38% und im Fall b) von 48,8% auf 2,33%. Der Unterschied zwischen den beiden Auswahlätzen ist bei kleinen Grundgesamtheiten am größten, z. B. rd. 46% bei einer Grundgesamtheit von 500 Einheiten, nur noch 2,1% bei einer Grundgesamtheit von 3 000 Einheiten und nur 0,05% bei einer Grundgesamtheit von 20 000 Einheiten. Das heißt, von einer gewissen Größenordnung der Grundgesamtheit ab wird er praktisch bedeutungslos.

Für die Lage dieser Grenze gibt es kein absolutes Kriterium. In der Stichprobenpraxis wird die Berücksichtigung des Endlichkeitsfaktors in der Regel als geboten betrachtet, wenn der Stichprobenumfang 5% oder mehr des Umfangs der Grundgesamtheit ausmacht¹⁾.

Nunmehr bleibt zu untersuchen, wie sich eine Veränderung der Anteile der fehlerfreien bzw. fehlerhaften Einheiten und damit die Größe der Varianz bei sonst gleichen Bedingungen auf den Stichprobenumfang auswirkt.

Die Varianz PQ ist offensichtlich am größten, wenn $P = Q = 0,5$, sie beträgt dann $PQ = 0,25$. Bei jeder anderen Kombination ist die Varianz kleiner (z. B. oben $PQ = 0,95 \cdot 0,05 = 0,0475$). Das bedeutet, daß bei im übrigen unveränderten Erfordernissen (z. B. N unbegrenzt groß; $e = 0,02$; $t = 2$) der Stichprobenumfang um so kleiner sein kann, je mehr sich die Größenordnungen der beiden Anteile voneinander unterscheiden (z. B. gilt unter den genannten Bedingungen bei

¹⁾ Vgl. Kellerer (1953), S. 37 f.; anderer Ansicht z. B. Dodge/Romig (a. a. O.), die die Grenze erst bei $n > 0,1 N$ ziehen (s. S. 26).

$P = Q = 0,5$ die Forderung $n = 2\,500$ und bei $PQ = 0,95 \cdot 0,05 = 0,0475$ die Bedingung $n = 475$). Denn je geringer der Fehleranteil in der Grundgesamtheit ist, desto *gedrungener* ist die Kurve, die die Häufigkeitsverteilung der Fehleranteile q der Stichproben umschließt¹⁾.

Bei konstantem N und n ist aber die Verteilung desto asymmetrischer, „*schiefer*“, je kleiner Q . Man muß daher auch die Frage untersuchen, inwieweit man sich bei der *Schätzung kleiner Anteile* noch der *Gaußschen* Näherungsbeziehung bedienen kann²⁾ und welche Auswirkungen die Begrenzungen des Anwendungsbereichs gegebenenfalls für die Festlegung der Bestimmungsfaktoren in den Planungsformeln (24) bzw. (25) haben.

Die *Gaußsche Näherungsbeziehung* kann dann *nicht mehr zur Grundlage von Anteilsschätzungen* gemacht werden, wenn die Anteile außerhalb des Bereiches

$$(27) \quad \frac{t^2}{n + t^2} \leq P \leq \frac{n}{n + t^2} \quad \text{bzw.} \quad \frac{n}{n + t^2} \geq Q \geq \frac{t^2}{n + t^2}$$

liegen³⁾.

Für die Beurteilung der Auswirkungen dieser Eingrenzungen auf die Planungsformeln erscheint es vorteilhaft, von der Relation

$$(28) \quad Q \geq \frac{t^2}{n + t^2}$$

auszugehen. Löst man sie nach n auf, dann erhält man

$$(29) \quad n \geq \frac{t^2 P}{Q}$$

während der erforderliche Mindeststichprobenumfang gemäß (18) und (23)

$$(30) \quad n = \frac{t^2 P Q}{e^2}$$

betragen muß. Man kann daher Formel (24) bzw. (30) nur anwenden, wenn

$$(31) \quad \frac{t^2 P}{Q} \leq \frac{t^2 P Q}{e^2}$$

woraus nach Kürzung und Umformung für den Fall $Q \leq P$ folgt

$$(32) \quad Q - e \geq 0$$

Um die *Gaußsche Näherungsbeziehung* einer Anteilsschätzung zugrunde legen zu können, darf also der *Meßwert e für den Vertrauensbereich nicht größer angesetzt werden als der Fehleranteil Q* . Das heißt, daß, wenn Q sehr klein ist, auch e sehr

¹⁾ Bei extrem klein werdenden Fehleranteilen kann sich diese Tatsache auf die Beziehung zwischen t und $F(t)$ auswirken, s. *Dinter*, a. a. O., S. 90 f.

²⁾ Zur Frage der Zweckmäßigkeit der Verwendung einer *Poisson*-Verteilung, einer *Binomial*-Verteilung oder einer *Normalverteilung* als angemessene Approximation für die tatsächliche Häufigkeitsverteilung der Schätzwerte siehe *Schaafsma*, A. H., *Willemze*, E. G., *Moderne Qualitätskontrolle*, 2. Auflage, Eindhoven 1959, S. 401 ff., ferner *Stringer*, K. W., *Some Basic Concepts of Statistical Sampling in Auditing*, JofA, November 1961, S. 63 ff.

³⁾ *Anderson*, a. a. O., S. 110 und insbesondere S. 254. Die dort benutzten Symbole p und q entsprechen den hier benutzten Größen P und Q .

klein sein muß. Das erfordert dann allerdings einen erheblichen Stichprobenumfang, um bei gegebenem Sicherheitsgrad allein die Bedingung $e = Q$ zu erfüllen.

Einige *Beispiele* für den erforderlichen absoluten Mindestumfang der Stichprobe im Fall $e = Q$ sind in der folgenden Tabelle 8 zusammengestellt. Sie wurden an Hand der Formeln (24) bzw. (25) berechnet.

Tabelle 8: Minimalumfang der Stichprobe zur Schätzung extrem niedriger Anteile Q bei Vernachlässigung bzw. bei Berücksichtigung des Endlichkeitsfaktors unter den Bedingungen $e = Q$ und $t = 2$.

Anteile	P Q	0,98 0,02	0,99 0,01	0,999 0,001	
Stichprobenumfang					
a) bei Vernachlässigung des Faktors		196	396	3 996	
b) bei Berücksichtigung des Faktors		Einheiten	in % v. N	Einheiten	in % v. N
bei einer Grundgesamtheit von N = 500 Einheiten		141	28	222	44
1 000		164	16	284	28
5 000		189	4	367	7
10 000		193	2	381	4
100 000		196	0,2	395	0,4
				800	80
				2 222	44
				2 856	29
				3 843	4

Bei *sehr geringen Anteilen* von fehlerhaften Einheiten steigt der Mindestumfang der Stichprobe also wegen der starken Einengung des Vertrauensbereichs sowohl bei Berücksichtigung wie bei Vernachlässigung des Endlichkeitsfaktors mit abnehmendem Anteil der fehlerhaften Einheiten an. Daraus folgt, daß Stichprobenverfahren bei einem im Verhältnis zum Umfang der Grundgesamtheit extrem kleinen Fehleranteil wegen der notwendigen außerordentlichen relativen Größe der Stichprobe ihren Wert weitgehend — unter Umständen sogar ganz — einbüßen. Mit zunehmender Schiefe der Verteilung *nähert sich der erforderliche Stichprobenumfang solchen Größenordnungen, die auch für Entdeckungsstichproben benötigt werden* (vgl. S. 22).

Weiß man von organisatorisch *untergliederten Grundgesamtheiten*, daß ihre Teilgesamtheiten unterschiedliche Fehleranteile aufweisen oder lassen sich Teilgesamtheiten mit unterschiedlichen Fehleranteilen zu Auswahlzwecken abgrenzen, so kann man u. a. die auf S. 42 dargestellten mathematischen Beziehungen (Formel 20) dazu benutzen, einen Plan für ein wirksameres *Auswahlverfahren* als die uneingeschränkte Zufallsauswahl zu entwerfen, also ein solches, bei dem man unter Konstanthalten der Varianz der Stichprobenergebnisse mit einem unter Umständen beträchtlich geringeren Stichprobenumfang auskommt, als er bei uneingeschränkter Zufallsauswahl benötigt wird. Stichprobenverfahren mit mehr als zwei Auswahlstufen sind nach entsprechend komplizierteren Formeln zu planen.

Von mehreren alternativen zweistufigen Auswahlplänen zur Untersuchung einer Grundgesamtheit mit gegebener Struktur muß derjenige als der beste angesehen werden, bei dem es gelingt, jeden der beiden Summanden in Formel (20) möglichst

gering zu halten. Die Formel zeigt, daß z. B. jede Vergrößerung der Zahl n' der erststufigen Stichprobeneinheiten beide Summanden auf der rechten Seite der Gleichung verkleinert. Eine Vergrößerung der Stichprobenumfänge n_i der zweiten Auswahlstufe verkleinert bei unveränderter Zahl der ausgewählten Primäreinheiten nur den zweiten Summanden. Bei gegebener Gesamtzahl der auszuwählenden Sekundäreinheiten werden also am besten möglichst viele Primäreinheiten und innerhalb dieser jeweils nur ein kleiner Teil von Sekundäreinheiten ausgewählt. Ferner läßt sich aus Formel (20) erkennen, daß bei großer Varianz der Primäreinheiten und kleinen Varianzen unter den Sekundäreinheiten der einzelnen Primäreinheiten möglichst ein geschichtetes Verfahren anzuwenden ist. Umgekehrt lassen hohe Varianzen bei den Sekundäreinheiten der einzelnen Primäreinheiten bei gleichzeitiger minimaler Varianz zwischen den Primäreinheiten die Klumpenauswahl als das geeignete Verfahren erscheinen. Quantitative Aussagen über den Rationalisierungseffekt eines bestimmten Verfahrens gegenüber der uneingeschränkten Zufallsauswahl sind nur bei Kenntnis der jeweiligen Struktur der Grundgesamtheit, also nur von Fall zu Fall möglich.

Mehrphasige Auswahlverfahren haben bei Schätzstichproben entweder das Ziel, in den ersten Phasen Anhaltswerte für die Größenordnung der interessierenden Merkmalwerte, also z. B. den Fehleranteil und die Art der Verteilung gleichartiger Merkmalwerte, zu gewinnen, um die Hauptphase des Verfahrens zweckentsprechend planen zu können, oder sie sollen die zunächst unzureichende Genauigkeit des Stichprobenergebnisses durch Vergrößerung des Stichprobenumfangs erhöhen. Auch hier sind generelle quantitative Aussagen bezüglich des damit verbundenen Rationalisierungseffekts nicht möglich (anders als im Fall der Annahmestichproben, s. S. 34).

Schließlich hängt der Stichprobenumfang auch noch vom erforderlichen *Sicherheitsgrad* und *Vertrauensbereich* ab. Allgemein kann festgestellt werden, daß bei gegebener Varianz in der Grundgesamtheit der erforderliche Stichprobenumfang mit dem Quadrat von t wächst und andererseits mit wachsendem Quadrat von e fällt. Schreibt der Prüfer also z. B. einen *hohen Sicherheitsgrad und einen engen Vertrauensbereich vor*, so wird selbst bei *kleiner Varianz in der Grundgesamtheit ein erheblicher Stichprobenumfang erforderlich*.

Die Abhängigkeit der Varianz vom Fehleranteil erlaubt es, die bei verschiedenen Größen des Fehleranteils und des Umfangs der Grundgesamtheit für verschiedene Genauigkeitserfordernisse (Sicherheitsgrad und Vertrauensbereich) *erforderlichen Stichprobenumfänge zu tabellieren*¹⁾.

c) Anwendbarkeit auf das Prüfungsgebiet

(1) Prüferische Auswertung

Zusammenfassend läßt sich aufgrund der vorstehenden Untersuchungen feststellen, daß Stichproben für Anteilsschätzungen bei gegebenem Sicherheitsgrad und Vertrauensbereich um so rentabler sind, 1. je größer die Grundgesamtheit ist und 2. je mehr sich die Größenordnungen der fehlerfreien und fehlerhaften Anteile voneinander unterscheiden. Bei extrem kleinem Fehleranteil muß der in die

¹⁾ Z. B. Brown, R. J./Vance, L. L., Sampling Tables for Estimating Error Rates or Other Proportions, Berkeley 1961.

Formel einzusetzende Vertrauensbereich sehr klein sein — weil $(Q - e)$ nicht negativ werden darf — mit der Folge, daß in solchen Fällen der Stichprobenumfang sehr groß wird, so daß Vollerhebungen zweckmäßig werden können.

Bei der Prüfung der formellen Ordnungsvorschriften im Rahmen der Abschlußprüfung lassen sich also Schätzstichproben für Anteilsschätzungen rationell verwenden, wenn die Grundgesamtheit groß genug ist, was bereits bei Mittelbetrieben der Fall sein kann, und der Anteil fehlerhafter Einheiten in der Grundgesamtheit nicht übermäßig groß, aber auch nicht extrem klein ist. In der Regel werden hohe Prozentzahlen fehlerhafter Einheiten nur sehr selten anzutreffen sein. Wo dies doch der Fall ist, genügt bereits diese Erkenntnis, um prüferische Folgerungen daraus zu ziehen. Umgekehrt werden extrem kleine Fehleranteile zweckmäßigerweise nicht durch Schätzstichproben festgestellt, wenn es auf große Aussagesicherheit oder möglichst absolute Fehlerfreiheit ankommt. In diesen Fällen müssen entweder Vollerhebungen vorgenommen oder entsprechende organisatorische Maßnahmen veranlaßt werden, z. B. die Feststellung derselben Zahlen an zwei unabhängig voneinander arbeitenden Stellen (Führung von Gegenbüchern) und ihre regelmäßige Abstimmung. Es bleiben dann für die Anwendung von Stichproben die Prüfungen der Einhaltung derjenigen formellen Vorschriften, die zwar in der Regel beachtet, in einem gewissen Umfang aber auch nicht beachtet werden. Auf diesen Gebieten erlaubt die Stichprobentheorie die Feststellung des Anteils fehlerhafter Einheiten und damit des Grades der Wirksamkeit der Ordnungsvorschriften mit rationellem, d. h. nicht zu großem, Aufwand und mit bestimmbarer Sicherheit innerhalb vorgegebener Fehlergrenzen. Ferner ist festzustellen, daß die Theorie genügend Möglichkeiten bietet, das Auswahlverfahren der Organisation und der Struktur der Grundgesamtheit anzupassen und dadurch eine rationelle Erhebung zu planen.

Schließlich ist prüferisch von Bedeutung, daß relativ kleine Veränderungen des Stichprobenumfangs praktisch keine Wirkung auf die Größe von Sicherheitsgrad und Vertrauensbereich haben und daher zwecklos sind.

(2) Grundlagen für die Festsetzung der Bestimmungsfaktoren

Umfang und Fehleranteil der Grundgesamtheit

Entsprechend der eingangs geschilderten Zielsetzung werden die *Grundgesamtheiten* durch den Geltungsbereich der einzelnen Ordnungsvorschriften bestimmt. Zahl und Art dieser Grundgesamtheiten sind von Unternehmen zu Unternehmen verschieden. Da die Nachprüfung der Einhaltung der formellen Ordnungsvorschriften nur dann möglich ist, wenn sie in irgendeiner Form auf bestimmten Dokumenten durch Unterschrift oder Zeichen ausdrücklich bestätigt wird, werden automatisch die für die Aufnahme dieser Unterschriften und Zeichen vorgesehenen Dokumente zu den *Untersuchungseinheiten*. Die einzelnen *Untersuchungsmerkmale* ergeben sich aus den gestellten Prüfungsfragen, wie z. B.: Ist die Verbuchung der Geschäftsvorfälle genehmigt? Ist der verbuchte Beleg so entwertet, daß eine Wiederverwendung unmöglich ist? Ist die Bankabhebung oder -überweisung durch doppelte Unterschrift angewiesen?

Aufgrund der obigen Darlegungen (s. S. 44 ff.) kann festgestellt werden, daß nur bei verhältnismäßig *kleinen Grundgesamtheiten* ihr *Umfang* genau bestimmt oder, wenn er nur ungefähr bekannt ist, der größte zu erwartende Umfang zur Grund-

lage der Planung gemacht werden muß, wenn der Endlichkeitsfaktor berücksichtigt werden soll. In diesen Fällen besteht sonst die Gefahr, daß wegen des zu geringen Stichprobenumfangs die Genauigkeit der Ergebnisse den Anforderungen nicht entspricht.

Bei der vorläufigen Abschätzung der tatsächlichen *Größe des Fehleranteils* in der Grundgesamtheit muß bedacht werden, daß ein näher an der 50%-Marke liegender Fehleranteil nur mit entsprechend größerem Stichprobenumfang hinreichend genau bestimmt werden kann.

Für die *Planung der Stichprobe* folgt daraus, daß man, falls bereits Anhaltspunkte für die Größe des *gesuchten Anteils* vorliegen und sie nur wenig voneinander abweichen, zweckmäßig den Wert als Bestimmungsgröße festsetzt, der dem Wert 0,5 am nächsten kommt¹⁾. Denn mit *einer* Auswahlphase eine etwas größere Ausagesicherheit (Sicherheitsgrad) als die geforderte zu erhalten, ist rationeller, als diese Sicherheit erst nachträglich mit Hilfe einer unter Umständen nur kleinen zweiten Stichprobenphase anstreben zu müssen. Liegen jedoch keine Informationen vor oder weichen die vorhandenen stark voneinander ab, so ist es ratsam, in einer kleinen Vorstichprobe von 50 bis 100 Einheiten eine erste grobe Schätzung vorzunehmen oder aber den niedrigsten der bekannten Werte anzusetzen und genügend Ergänzungsphasen bei der Planung des Auswahlverfahrens zu berücksichtigen.

Deuten Anhaltspunkte oder die Voruntersuchung darauf hin, daß *extrem kleine Fehleranteile* zu erwarten sind, so muß der Prüfer entscheiden, ob der ursprünglich vorgesehene Sicherheitsgrad erforderlich ist und im bejahenden Fall, ob nicht die Vollerhebung oder geeignete organisatorische Maßnahmen rationeller sind als die Durchführung der — wie gezeigt umfangreichen — Stichprobenerhebung.

Die Untersuchung der Grundgesamtheit zur vorläufigen Abschätzung der Anteilgröße erschöpft sich jedoch im allgemeinen nicht mit der überschläglichen *Einschätzung* eines durchschnittlichen Fehleranteils für die ganze Grundgesamtheit. Es sollte vielmehr, insbesondere bei sehr großen und kompliziert organisierten Grundgesamtheiten, ermittelt werden, ob und welche *Unterschiede zwischen den Fehleranteilen in den einzelnen organisatorisch unterschiedenen Teilen der Grundgesamtheit* bestehen. Die hierbei erworbenen Kenntnisse über die organisatorische Gliederung der Grundgesamtheit und die Struktur ihrer Teile sind dann zu verwenden, um ein Auswahlverfahren zu planen, mit dem die höchste Wirksamkeit der Erhebung erzielt wird.

Auswahlgrundlagen und Auswahlverfahren

Als *Auswahlgrundlagen* für die Stichprobenauswahl können die bei der Untersuchung des innerbetrieblichen Kontrollsystems erworbenen Kenntnisse, insbesondere des Arbeitsablauf-Schaubildes, verwandt werden.

Es setzt die an den einzelnen Funktionen beteiligten Personen und die bei der Ausübung einer Funktion zu Grunde liegenden Unterlagen zueinander in Beziehung, gibt aber auch den Ablageort der entsprechenden Dokumente an. Dabei ist es von Bedeutung, daß das Schaubild auch *Hinweise auf die Abwicklung*

¹⁾ Liegt nur ein Anhaltspunkt vor, so wird natürlich dieser der Festsetzung zugrunde gelegt.

der Ausnahmen unter den Geschäftsvorfällen der Rechnungsperiode enthält und zeigt, wie diese Unterlagen geordnet und wo sie aufbewahrt sind.

Der Statistiker hat nun ein *Auswahlverfahren* zu planen, das den unterschiedlichen oder gleichartigen Anteilgrößen der Einheiten mit dem interessierenden Merkmal in den durch die Auswahlgrundlagen abgegrenzten Teilgesamtheiten möglichst gut angepaßt ist. Er kann dabei je nach den im praktischen Einzelfall vorliegenden Verhältnissen und entsprechend den Untersuchungszielen (ob z. B. nur ein Gesamtergebnis oder ob auch Detailergebnisse interessieren) entweder Teilgesamtheiten mit ähnlichen Anteilgrößen zu Schichten oder Teilgesamtheiten mit sehr unterschiedlichen Anteilgrößen zu Klumpen zusammenfassen.

Zum Beispiel werden bei der Prüfung der Kassenbelege auf die Ordnungsmäßigkeit ihrer Zahlungsanweisung *Schichten* nach der Größenordnung der Geldbeträge gebildet, nämlich eine solche, die alle Belege über Beträge unter DM 1 000,— umfaßt, für die nur eine Unterschrift eines Sachbearbeiters erforderlich ist, eine zweite, die alle Belege von Beträgen von DM 1 000,— bis unter DM 5 000,— enthält, bei denen die Unterschriften eines Sachbearbeiters und eines Gruppenleiters verlangt werden, und eine dritte, die die Belege von Beträgen ab DM 5 000,— aufwärts einschließt, bei denen die Anweisung durch einen Gruppenleiter und einen Prokuristen erforderlich ist. Aus diesen Schichten werden die Stichprobeneinheiten durch Zufallsauswahl bestimmt. Man kann die Kassenbelege auch nach dem Datum ihrer Ausstellung ordnen und wochenweise zusammenfassen und innerhalb einzelner durch Zufallsauswahl bestimmter Wochen alle Belege auf die vorgenannten Untersuchungsmerkmale prüfen (*Klumpenauswahl*). Wenn sich die Grundgesamtheit auf eine größere Zahl von Filialen verteilt, kann in einer ersten Auswahlstufe durch Zufallsauswahl eine Anzahl von Filialen ausgesucht und auf der zweiten Stufe anschließend bei diesen Filialen wie vorstehend beschrieben vorgegangen werden (*zweistufige Auswahl*).

In diesem Zusammenhang ist dann eine *Reihe von statistischen Problemen* zu lösen, wie etwa das der Festlegung der Schichtungs-, Anordnungs- oder Stufungsmerkmale, der Bestimmung der Anzahl der Schichten oder Stufen oder des Auswahlabstandes, der Abgrenzung der Schichten, des Stichprobenumfangs innerhalb der einzelnen Schichten oder Stufen, die Frage, ob und wie Schichtungen oder Anordnungen in den einzelnen Auswahlstufen vorgenommen werden sollen u. v. a. m.¹⁾ Stellt sich heraus, daß die tatsächliche Anteilgröße in einer oder mehreren der Auswahlgesamtheiten (Schichten, Stufen etc.) näher am 50%-Wert liegt, als die zunächst vermutete Anteilgröße in dieser Teilgesamtheit, so kann durch die Vergrößerung des Stichprobenumfangs in einer zusätzlichen Auswahlphase die Berichtigung vorgenommen werden.

Ein praktisches *Beispiel für die schrittweise Anpassung des Auswahlverfahrens an die tatsächlichen Verhältnisse* durch Ausnutzung der wachsenden Kenntnisse von der Struktur der Grundgesamtheit wird bei der Erörterung des Problems der stichprobenweisen Prüfung der Inventur an entsprechender Stelle wiedergegeben.

¹⁾ Statistisches Bundesamt (Hrsg.), *Stichproben in der amtlichen Statistik*, Stuttgart und Mainz 1960; Kellerer, H., *Theorie und Technik des Stichprobenverfahrens*, a. a. O.; Hansen, M. H./Hurwitz, W. N./Madow, W. G., *Sample Survey Methods and Theory*, Bd. I u. II., 3. Aufl., New York—London 1960; Deming, W. E., *Some Theory of Sampling*, New York—London 1950; ders., *Sample Design in Business Research*, New York—London 1960, u. v. a., vgl. auch *Literaturverzeichnis*.

Sicherheitsgrad und Vertrauensbereich

Über die Festsetzung der bereits besprochenen Faktoren hinaus bedarf es auch der Festsetzung des *Sicherheitsgrads* (*t*) und des *Vertrauensbereichs* (*e*) entsprechend den prüferischen Bedingungen. Hierfür sind allerdings objektive Bestimmungsfaktoren noch nicht erkennbar, so daß über die Feststellung von *Pougin*¹⁾, daß es auf das Ermessen des Abschlußprüfers ankommt, insoweit noch nicht hinausgegangen werden kann. Es ist jedoch denkbar, daß diese Bedingungen z. B. im Rahmen der hausinternen Übung und Erfahrung des Abschlußprüfers²⁾ oder durch Berufsnormen der Wirtschaftsprüfer festgelegt werden, wie es die Union Européenne des Experts Comptables Economiques et Financiers (UEC) aufgrund der Beratungen ihrer Commission des Mathematiques Financières gefordert hat³⁾.

3. Bedeutung der Ergebnisse der Prüfung der Sicherheit der Rechnungslegung für die Prüfung der materiellen Richtigkeit der Rechnungslegung.

Die Prüfung der Doppik wie der formellen Ordnungsvorschriften, sowohl der gesetzlichen wie der unternehmensinternen, vermittelt dem Prüfer ein Bild von der Qualität der Buchführung, nämlich ob sie zuverlässig und ordnungsmäßig geführt ist oder nicht, m. a. W. in welchem Maße er sich auf diese Aufzeichnungen verlassen kann. Aufgrund dieser Prüfungsergebnisse kann er sich dann entscheiden, in welchem Umfang er die Prüfung der materiellen Richtigkeit der Rechnungslegung ansetzen muß.

Zu dieser Frage hat sich *Pougin* in bezug auf die Bedeutung der Wirksamkeit des innerbetrieblichen Kontrollsystems geäußert; seine Ausführungen gelten aber allgemein für die Einhaltung der Ordnungsvorschriften. Er schreibt⁴⁾:

„Im Laufe der Untersuchung des Kontrollsystems kommt der Punkt, wo der Prüfer entscheiden muß, welchen Einfluß das Ergebnis der Untersuchung auf Verfahren und Umfang der eigentlichen Abschlußprüfung (d. h. also der materiellen Abschlußprüfung im Sinn der in dieser Arbeit benutzten Definition, Anm. d. Verf.) haben soll. Diese Entscheidung ist besonders schwerwiegend, wenn der Prüfer Mängel im Kontrollsystem aufgedeckt hat. Über ihre Auswirkung auf den Prüfungsumfang sind die Meinungen der amerikanischen Wirtschaftsprüfer geteilt: Die einen wollen die gesamte Prüfung ausdehnen, die anderen wollen den Prüfungsumfang nur dort erweitern, wo Mängel im Kontrollsystem gefunden worden sind.

Wo Schwächen im internen Kontrollsystem gefunden werden, muß der Prüfungsumfang in dem Maße ausgedehnt werden, wie es das Ermessen des Prüfers verlangt. Hierfür läßt sich keine Regel aufstellen, denn keine Situation ist einer anderen gleich. Die Prüfung mag 50 bis 70% des Prüfungsstoffes und in gewissen Fällen noch mehr umfassen.“

1) a. a. O., S. 95 f.

2) S. dazu u. a. *Brown, R.G.*, Statistical Sampling Tables for Auditors, JofA, Mai 1961, S. 49.

3) S. dazu: Unternehmen und Unternehmensberater im Gemeinsamen Europäischen Markt, Bericht über den Dritten Kongreß der UEC in Nizza 1958, hrsg. vom Institut der Wirtschaftsprüfer in Deutschland, Düsseldorf 1960, S. 72.

4) a. a. O., S. 94.

Auch andere deutsche Autoren wollen die Organisation der Buchführung, d. h. die Art des Buchführungssystems, also den Buchungsgang und die Art und Zahl der hierdurch gegebenen oder fehlenden zwangsläufigen Kontrollen bei der Planung der Abschlußprüfung berücksichtigt wissen, so *Hintner, Hertlein/Meisner, Adler/Düring/Schmaltz*¹⁾ u. a. Ähnlich wird in der amerikanischen Literatur zu diesem Thema Stellung genommen²⁾.

Nun erlaubt, wie die vorstehenden Untersuchungen zeigen, die Anwendung der Stichprobenverfahren in vielen Fällen eine rationelle Feststellung der Fehleranteile und damit eine Aussage über die Wirksamkeit der Kontrollleinrichtungen. Allerdings sind die verschiedenen Fehlerarten und Fehlerorte von unterschiedlicher Bedeutung für die Sicherheit der Rechnungslegung. Ein objektiver Maßstab für ihr Gewicht im Rahmen der gesamten Rechnungslegung ist bisher nicht gefunden worden, so daß eine Gesamtwertung der Qualität der Rechnungslegung bisher objektiv nicht möglich ist.

Wohl aber besteht mit Hilfe der Stichprobenmethoden die Möglichkeit, den Umfang der verschiedenen Fehlerarten an den verschiedenen Fehlerorten und damit die Wirksamkeit der einzelnen Ordnungs- und Kontrollvorschriften objektiv zu messen und so ihre Wirksamkeit in verschiedenen Prüfungszeiträumen oder bei verschiedenen Abteilungen (z. B. Filialen) oder Unternehmen zu vergleichen.

Das genügt aber bereits, um die Wirksamkeit einer bestimmten Kontrollmaßnahme für ein bestimmtes Buchführungsgebiet zu bestimmen.

So kann z. B. aus der Wirksamkeit der Kontrollvorschriften für den Kontokorrentverkehr ein Schluß auf die Zuverlässigkeit dieses Buchführungsgebiets gezogen werden. Bestehen für ein solches Teilgebiet mehrere Kontrollvorschriften, so kann das von *Mackenzie* erwähnte Verfahren (s. S. 37) Anwendung finden.

Ist es also nicht möglich, eine Gesamtbewertung der Sicherheit der Rechnungslegung an Hand der Feststellung der Zahl der fehlerhaften Buchungsvorgänge durchzuführen, so kann dies doch sinnvoll für Teilgebiete geschehen. Damit kann aus den Ergebnissen der formellen Prüfung dieser Teilgebiete ein Maß für die Bestimmung des Umfangs der materiellen Prüfungen dieser Teilgebiete abgeleitet werden.

Diese Zusammenhänge hat *A. Ridilla*³⁾ in einem Aufsatz „A Simplified Statistical Technique for Use in Verifying Accounts Receivable“ — ebenfalls nur für ein bestimmtes Teilgebiet der Rechnungslegung, nämlich für eine Debitorenbuchführung — untersucht. Er schreibt:

„Der Sicherheitsgrad (degree of accuracy), der von der Stichprobe verlangt wird, ist direkt abhängig von der Bewertung des innerbetrieblichen Kontrollsystems der Debitoren des Mandanten durch den Prüfer. Wenn er die innerbetriebliche

1) *Hintner*, a. a. O., S. 68; *Hertlein, A./Meisner, K.*, Abschluß und Prüfung der Unternehmungen, 4. Aufl., Wiesbaden 1956, S. 94 f.; *Adler/Düring/Schmaltz*, a. a. O., § 135 an mehreren Stellen.

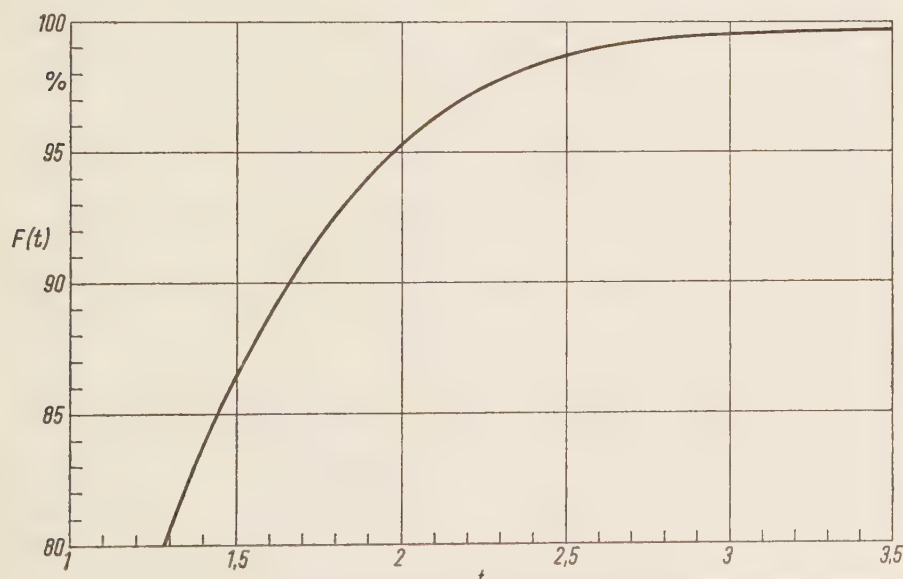
2) American Institute of Certified Public Accountants, Generally Accepted Auditing Standards, 1954; *Brink, V. Z.*, The Independent Auditor's Review of Internal Control, *JofA*, Bd. 73, 1942, S. 430; *Sprague, W. D.*, Testing the Adequacy of Internal Control, *JofA*, Bd. 101, 1956, S. 49 ff.

3) „The Accounting Review“, Oktober 1959, S. 548.

Kontrolle schwach oder schlechter als wünschenswert ansieht, sollte er auf einem hohen Sicherheitsgrad seiner Hauptstichprobe bestehen. Dies bedeutet, daß der Prüfer in seinen Anforderungen an den Echtheitsbeweis sowohl der Kundenkonten als des Debitoren-Sachkontos strenger sein muß, sobald die innerbetriebliche Kontrolle schwach ist. Wenn die innerbetriebliche Kontrolle außerordentlich gut erscheint, kann er es vertreten, seine Stichprobenprüfungen oder den von der Stichprobe verlangten Sicherheitsgrad zu verringern.“

Ridilla zeigt dann eine im Prinzip bereits auf Seite 43 dargestellte Tabelle, die den Zusammenhang zwischen dem gewünschten Sicherheitsgrad $F(t)$ für die materielle Prüfung und dem Wert t bei Normalverteilung der Merkmale darstellt. Obwohl diese Tabelle nur ein Ausschnitt ist, der Sicherheitsgrade von 80,0% bis 99,95% enthält, sollte sie — nach *Ridilla* — genügen, „um Differenzen in der Wirksamkeit der innerbetrieblichen Kontrolle von einer Organisation zur anderen in Betracht zu ziehen“. Er führt dann aus, daß bei einem ausgezeichneten Stand des innerbetrieblichen Kontrollsystems ein Sicherheitsgrad von 80% für die materielle Stichprobenschätzung genügt, was einem t -Wert in der Formel von 1,28 entspricht, und daß bei einem sehr schwachen innerbetrieblichen Kontrollsystem ein höherer Sicherheitsgrad für die materiellen Schätzungen, und zwar bis zu 99,95% verlangt werden muß, was einem $t = 3,5$ entspricht.

Zur Veranschaulichung der Zusammenhänge diene die folgende Figur 4 (vgl. Tabelle 5).



Figur 4: Beziehung zwischen Sicherheitsgrad $F(t)$ und t im Fall einer Normalverteilung der Merkmalwerte

Allerdings wird auch von *Ridilla* der Zusammenhang zwischen der Wirksamkeit der innerbetrieblichen Kontrollmaßnahmen (ausgezeichnet — sehr schwach) und dem für die materielle Prüfung der Debitoren zu verlangenden Sicherheitsgrad (80% bis 99,95%) nicht aufgrund fester Maßstäbe hergestellt. Mit Stichproben-

erhebungen läßt sich zwar die Wirksamkeit der innerbetrieblichen Kontrollmaßnahmen, wie vorstehend dargelegt, durch den Prozentanteil der festgestellten Fehler messen. Aber auch wenn dieses Maß gefunden ist, fehlt noch ein objektiver Zusammenhang zwischen dem Qualitätswert der innerbetrieblichen Kontrollmaßnahmen und dem für die materiellen Prüfungsstichproben zu verlangenden Sicherheitsgrad.

IV. Prüfung der materiellen Richtigkeit der Rechnungslegung

1. Prüfung der Bestände

A) Gesamtwерtermittlungen

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Die materielle Richtigkeit des Jahresabschlusses erfordert, daß alle Bestandsposten nach Art und Menge inventarisiert, d. h. erfaßt und entsprechend den gesetzlichen Vorschriften bewertet werden.

Aus der Zweiteilung der Inventur in mengenmäßige Erfassung der Bestände und ihre anschließende Bewertung ergeben sich drei Möglichkeiten für ihre Durchführung:

- a) lückenlose Aufnahme und lückenlose Bewertung (Vollerhebung)
- b) lückenlose Aufnahme und stichprobenweise Bewertung (Stichprobenaufbereitung)
- c) stichprobenweise Aufnahme und Bewertung (Stichprobenerhebung).

Die *Prüfung* des vom Unternehmen praktizierten Inventurverfahrens kann entweder durch Prüfung der Inventur, also der Aufnahmearbeiten, oder durch nachträgliche Prüfung der Inventurunterlagen erfolgen, der Prüfer kann aber auch selbst Bestände aufnehmen und seine Ergebnisse mit denen des Unternehmens vergleichen.

Die *vollständige Beobachtung* der Inventurarbeiten des Unternehmens begegnet häufig praktischen Schwierigkeiten, weil z. B. nicht genügend Prüfer vorhanden sind, um alle Inventurkolonnen während ihrer gesamten Inventurarbeiten zu überwachen. Andererseits ist es für den Prüfer unzumutbar, nur einen *Teil der Inventurarbeiten* persönlich zu beobachten, wenn er aus den beobachteten Teilgebieten auf den Wert der Gesamtheit schließen will. Die stichprobenweise Beobachtung einzelner Aufnahmeakte durch den Prüfer läßt erwarten, daß gerade diese Aufnahmen mit besonderer Sorgfalt durchgeführt werden und daher materiell genauer sind, als die nicht beobachteten. Wegen der größeren Genauigkeit dieser Aufnahmen wird sich dann das Schätzergebnis in der Regel nennenswert vom Ergebnis des Unternehmens unterscheiden, was zu neuen Untersuchungen, d. h. zu einer Ausdehnung des Stichprobenumfangs Veranlassung gibt, um die Ursachen der Abweichungen zu erkennen.

Statt einer stichprobenweisen Beobachtung der Inventurarbeiten leistet dann eine Prüfung, die feststellt, ob die *Inventurrichtlinien* zweckmäßig aufgestellt sind und ordnungsmäßig angewandt werden, bessere Dienste. Hinsichtlich der Einzelheiten der Prüfung von Organisations- und innerbetrieblichen Kontrollmaßnahmen kann auf die obigen Ausführungen verwiesen werden (s. S. 38 ff.).

Will der Prüfer die Bestände des Unternehmens aber *selbst* aufnehmen und sich nicht nur auf die Prüfung der Inventurunterlagen beschränken, dann sollte er unabhängig vom Inventurverfahren des Unternehmens einzelne stichprobenweise ausgesuchte Einheiten, die schon vom Unternehmen inventarisiert sind, selbständig, gegebenenfalls mit Hilfe des Personals des Unternehmens erfassen, um Verzerrungen der Ergebnisse durch die unterschiedliche Genauigkeit der vom Prüfer beobachteten und der nicht von ihm kontrollierten Bestandsaufnahmen zu vermeiden.

Es steht dem Prüfer dann grundsätzlich frei, ob er es bevorzugt, progressiv die Aufnahme stichprobenweise nachzuvollziehen oder retrograd einige Inventuraufzeichnungen auszuwählen und festzustellen, ob sie mit den Tatsachen übereinstimmen. Die *progressive* Prüfungsweise bietet die Chance, in den Aufzeichnungen nicht erfaßte, aber vorhandene Bestände aufzuspüren, der *retrograde* Prüfungsweg erlaubt die Aufdeckung von Scheinpositionen in den Listen, denen keine echten, d. h. tatsächlichen, Bestände zugrunde liegen.

Bei *lückenloser Aufnahme* und *lückenloser Bewertung* seitens des Unternehmens kann der Prüfer z. B. stichprobenweise die Aufnahme und lückenlos die Bewertung untersuchen. Er kann auch einige Bestandsposten zufallsmäßig herausgreifen und nur sie, aber sowohl ihre Aufnahme als auch ihre Bewertung, prüfen. Letzteres ist dem Prüfer auch dann möglich, wenn das Unternehmen seine Bestände *lückenlos aufgenommen* und *stichprobenweise bewertet* hat. Er ist bei der Auswahl seiner Stichprobeneinheiten dann nicht an die vom Unternehmen bewerteten Einheiten gebunden, sondern kann selbst Bewertungen durchführen und aus diesen Ergebnissen auf den Gesamtwert der Grundgesamtheit, der seine Stichproben jeweils entnommen sind, schließen. Eine prüferisch befriedigende Aussage erhält er, wie eingangs bereits festgestellt, immer nur für die Prüfungsgegenstände (Stichprobeneinheiten), die er im Sinne einer Wurzelprüfung voll untersucht, d. h. bei denen er sowohl die Aufnahme wie die Bewertung in allen Prüfungsmerkmalen prüft, also bei Erzen z. B. nicht nur Menge und Preis, sondern auch den Erzgehalt. Wo das Unternehmen bereits mit *Stichproben bei der Inventur* arbeitet, kann der Prüfer grundsätzlich die Stichproben des Unternehmens lückenlos oder stichprobenweise prüfen, im letztgenannten Fall also eine Unterstichprobe ziehen und aus ihr auf die Genauigkeit des Ergebnisses der Unternehmensinventur schließen. (Voraussetzung einer wirksamen materiellen Prüfung eines vom Unternehmen durchgeführten stichprobenweisen Inventurverfahrens ist eine sorgfältige statistische Überprüfung von Planung, Vorbereitung, Durchführung und Auswertung der Inventurstichproben durch den Prüfer.) Er kann ferner *unabhängig vom Unternehmen* eine Stichprobe aus der Grundgesamtheit ziehen. Er hat dann die Möglichkeit, die beiden Ergebnisse zu verknüpfen¹⁾.

¹⁾ Vgl. z. B. *Deming* (1960), S. 150 ff. In diesem, unter S. 65 ff., beschriebenen Fall der Stichprobeninventur von Halbfabrikaten kann der Prüfer selbst eine der beiden Unterstichproben ziehen oder ihre Durchführung voll überwachen.

Bei dem *prüferischen Ziel*, Menge und Wert der Bestände festzustellen, handelt es sich statistisch um die *Ermittlung der Gesamtwerte* von Grundgesamtheiten. Eine solche Aufgabe läßt sich ebenfalls mit *Schätzstichproben* lösen, und zwar mit solchen für die Gesamtwertermittlung.

b) Schätzung von Gesamtwerten mittels Schätzstichproben

(1) Formelapparat

Bei der Planung von Stichproben für die Schätzung von Gesamtwerten von Grundgesamtheiten kommt es — wie im Fall der Schätzung von Anteilen (s. S. 40 ff.) — ebenfalls darauf an, den Umfang der Stichprobe so festzulegen, daß bei dem geplanten Auswahlverfahren der Zufallsfehler mit vorgegebenem Sicherheitsgrad eine bestimmte Größe nicht überschreitet, m. a. W. daß mit bestimmtem Sicherheitsgrad eine Stichprobe ausgewählt wird, die einen Schätzwert für den Gesamtwert liefert, in dessen Vertrauensbereich (von vorgeschriebener Größe) der bei gleich sorgfältiger Vollerhebung sich ergebende Gesamtwert liegt.

Den Ausführungen zum Formelapparat bei den Stichproben für Anteilsschätzungen (s. S. 40 ff.) war bereits zu entnehmen, daß die *grundlegenden Beziehungen* zwischen dem Mittelwert aller bei einem bestimmten Auswahlplan gleich möglichen Schätzwerte und dem Ergebnis der gleich sorgfältigen Vollerhebung sowie zwischen der Varianz der Schätzwerte und der Varianz der Grundgesamtheit nicht nur im Fall der Anteilsschätzung, sondern auch bei den Schätzungen von Mittelwerten und Gesamtwerten gelten.

Die *Unterschiede im Formelapparat*, die durch die andere Zielsetzung der Erhebung, d. h. durch die andere Art des Schätzwerts (oben Anteilwert, jetzt Gesamtwert), verursacht sind, bestehen im wesentlichen darin, daß an die Stelle der Anteilwerte die Mittelwerte und an die Stelle des Produkts der Anteilwerte die Varianz der Merkmalwerte der Grundgesamtheit treten.

Bei *uneingeschränkter Zufallsauswahl* gilt, daß der Mittelwert $\bar{X}$ aller möglichen X gleich $N\mu$ ist.

$$(33) \quad \bar{X} = N\mu$$

Für die Varianz σ_X^2 der Schätzwerte X gilt bei Vernachlässigung des Endlichkeitsfaktors

$$(34) \quad \sigma_X^2 = N^2 \sigma_{\bar{X}}^2 = \frac{N^2 \sigma^2}{n}$$

bei Berücksichtigung des Endlichkeitsfaktors

$$(35) \quad \sigma_X^2 = N^2 \sigma_{\bar{X}}^2 = \frac{N^2 \sigma^2}{n} \frac{N - n}{N - 1}$$

Als *Schätzwert* für das in der Regel unbekannte σ^2 kann der Wert s^2 der Stichprobe genommen werden, wobei $s^2 = [1/(n - 1)] \sum_{i=1}^n (X_i - \bar{X})^2$; dieser Wert wird ebenfalls als Stichprobenvarianz bezeichnet [vgl. Kellerer (1953), S. 40 ff., 48 f.].

Im Fall eines *zweistufigen Auswahlverfahrens* ist die Varianz der Schätzwerte für den Gesamtwert¹⁾ — analog zu (20) —

$$(36) \quad \sigma_X^2 = \frac{N'^2 \sigma_e^2}{n'} \frac{N' - n'}{N' - 1} + \frac{N'}{n'} \sum_{i=1}^{N'} \frac{N_i^2 \sigma_i^2}{n_i} \frac{N_i - n_i}{N_i - 1}$$

wobei σ_e^2 die Varianz zwischen den N' Primäreinheiten und σ_i^2 die Varianz innerhalb der i . Primäreinheit sowie $\sum_{i=1}^{N'} N_i = N$ und $\sum_{i=1}^{N'} n_i = n$ sind. Auch hier lassen sich die geschichtete und die Klumpenstichprobe als Grenzfälle einer zweistufigen Auswahl auffassen (vgl. S. 42).

Als *Näherungswerte* für die unbekannten σ_e^2 und σ_i^2 sind deren Schätzwerte s_e^2 und s_i^2 zu nehmen.

Der Zusammenhang zwischen dem Streuungsmaß σ_X und den Häufigkeiten bestimmter Abweichungen ($X - N\mu$) bzw. dem *Sicherheitsgrad* (t), einen bestimmten *Vertrauensbereich* ($X \pm e$) einzuhalten, und der Größe e ist der gleiche wie der bereits auf S. 42 ff. beschriebene (s. Formel 23).

Zur Ermittlung des *benötigten Stichprobenumfangs* für Gesamtwertschätzungen ist in (23) an die Stelle von σ_Y^2 der Wert σ_X^2 gemäß (34) bzw. (35) zu setzen. Dann gilt bei *uneingeschränkter Zufallsauswahl* und bei Vernachlässigung des Endlichkeitsfaktors

$$(37) \quad n \geq \frac{t^2 N^2 \sigma^2}{e^2}$$

bei Berücksichtigung des Endlichkeitsfaktors

$$(38) \quad n \geq \frac{t^2 N^3 \sigma^2}{e^2 (N - 1) + t^2 N^2 \sigma^2}$$

(2) Wirkung der Bestimmungsfaktoren auf den Stichprobenumfang

Der *Umfang der Grundgesamtheit* kommt in Formel (37) durch den Faktor N^2 zur Geltung. Sein Auftreten erklärt sich daraus, daß der Wert e hier als Maß für den Vertrauensbereich für den Gesamtwert der Grundgesamtheit definiert ist.

In Formel (38) tritt der Umfang der Grundgesamtheit außerdem aufgrund der Berücksichtigung des Endlichkeitsfaktors in Erscheinung. Diesbezüglich kann auf die Ausführungen auf Seite 44 ff. verwiesen werden, denn wenn man den Stichprobenumfang in Formel (37) mit n_1 und denjenigen in Formel (38) mit n_2 bezeichnet, läßt sich zwischen ihnen die schon erwähnte Beziehung (26) ableiten.

Zur *Varianz* verhält sich der erforderliche Stichprobenumfang, wie die Formeln (37) und (38) zeigen, ebenso wie im Fall der Anteilsschätzung [s. Formel (24) und (25)], so daß bei großer Varianz auch ein großer Stichprobenumfang und unter sonst gleichen Bedingungen bei kleiner Varianz nur ein kleiner Stichprobenumfang erforderlich ist.

¹⁾ Vgl. Kellerer (1953), S. 145.

Eine allgemeingültige *obere Grenze* für die Varianz gibt es bei Schätzungen von Gesamtwerten und Mittelwerten nicht — im Gegensatz zum Fall der Anteilsschätzungen, wo die Varianz bekanntlich den Wert $PQ = 0,25$ nicht überschreiten kann —. Es gibt ferner auch keine Bedingung für den Höchstwert des Vertrauensbereichs, wie sie bei den Anteilsschätzungen durch $e \leq Q$ für $Q \leq P$ (vgl. Formel 32) für die Anwendbarkeit des Formelapparates besteht.

Hinsichtlich des Einflusses des *Auswahlverfahrens* auf den Stichprobenumfang gelten, wie ein Vergleich der Formeln (20) und (36) zeigt, bezüglich der geschichteten, der Klumpen- und der mehrstufigen Auswahlverfahren die gleichen Grundsätze, wie bei den Schätzstichproben zur Anteilsermittlung. Ebenso kann hinsichtlich der mehrphasigen Auswahlverfahren auf die obigen Ausführungen verwiesen werden (vgl. S. 49f.).

Darüber hinaus ist indessen bemerkenswert, daß insbesondere für die hier erörterten Schätzungen bereits Formeln aufgestellt worden sind, in denen die unterschiedlichen Erhebungskosten für die verschiedenen Schichten bzw. für die Auswahlseinheiten der verschiedenen Auswahlstufen sowie die Auswahlphasen berücksichtigt werden¹⁾.

Auf den Einfluß des Verhältnisses zwischen dem Faktor t für den *Sicherheitsgrad* und der Größe e für den *Vertrauensbereich* auf den erforderlichen Stichprobenumfang ist ebenfalls, und zwar auf Seite 49, bereits eingegangen worden.

c) Anwendbarkeit auf das Prüfungsgebiet

(1) Prüferische Auswertung

Die Bestimmungsfaktoren für die Stichproben zur Schätzung von Gesamtwerten sind, wie vorstehend erläutert, denen für die Stichproben zur Anteilsschätzung bis auf die Varianz und den Vertrauensbereich gleich. Die Unterschiede zwischen den Varianzen und den Vertrauensbereichen sind auf die verschiedene Art der Untersuchungsmerkmale (kategoriale-meßbare) und der Erhebungsziele (Anteilsschätzung — Gesamtwertschätzung) zurückzuführen. Der Einfluß der Bestimmungsfaktoren auf den Stichprobenumfang wird davon jedoch, wie ein Vergleich der Formeln (37) bzw. (38) mit den Formeln (24) bzw. (25) zeigt, nicht berührt. Infolgedessen kann auch hier grundsätzlich auf jene Ausführungen verwiesen werden (s. S. 50f.).

In der Prüfungspraxis sind die zu untersuchenden Grundgesamtheiten der Bestände von sehr unterschiedlicher Größe. Es gibt Unternehmen, deren Bestände sich nur auf wenige Posten verteilen, und solche mit sehr zahlreichen und verschiedenen Einzelposten. Sehr häufig läßt sich feststellen, daß geringwertige Posten, gleichgültig ob körperliche Bestände oder Forderungen, sehr zahlreich, insgesamt aber trotzdem nur mit verhältnismäßig kleinem Anteil am Gesamtwert vorhanden sind, während umgekehrt wenige relativ große Posten einen erheblichen Anteil am Gesamtwert der jeweiligen Bestandskategorie haben. Das drängt dann zu einer Schichtenbildung, wobei die Bestände mit zahlreichen Posten besonders rationell mit Stichprobenverfahren

¹⁾ Z. B. Kellerer (1953), S. 110 f., 148 f.; Vance/Neter, a. a. O., S. 218 ff., 226 f.; Deming (1960), an verschiedenen Stellen.

aufgenommen und bewertet werden können, während die schwach besetzten Schichten mit hohen Einzelwerten relativ stark, wenn nicht gar lückenlos, zu erfassen sind.

Eine exakte Bewertung ist oft nicht möglich (z. B. bei den Debitoren), andererseits aber auch nicht erforderlich, weil der Gesetzgeber dem Kaufmann einen nicht unerheblichen Bewertungsspielraum gelassen hat und deshalb alle Werte innerhalb dieses Spielraums als gesetzmäßig anzusehen sind. Entscheidend ist letztlich die Feststellung, ob der ausgewiesene Bilanzwert der Bestände (Menge $\times$ Einzelwert) innerhalb des gesetzlich zulässigen Bewertungsspielraums liegt und u. U. wie groß die Differenz zwischen dem ausgewiesenen Bilanzwert und dem gesetzlich erlaubten Grenzwert ist. Bleibt der Bilanzausweis unterhalb der gesetzlichen Bewertungsgrenze, so hat der Prüfer einen mit der Differenz dieser Werte wachsenden Ermessensspielraum, liegt der Ausweis oberhalb der Grenze, muß der Prüfer das Ausmaß der Überbewertung schätzen, um den Ausweis entsprechend berichtigen zu können. Mit Bezug auf die Passivbestände sind die Worte „oberhalb“ und „unterhalb“ zu vertauschen und „Unterbewertung“ statt „Überbewertung“ zu setzen.

Hieraus ergibt sich, daß es bei der Prüfung der Bestände genügt, ihre Werte näherungsweise zu bestimmen, wobei der Grad der erforderlichen Genauigkeit nach prüferischen Gesichtspunkten zu bestimmen ist.

(2) Grundlagen für die Festsetzung der Bestimmungsfaktoren

Umfang der Grundgesamtheit und Varianz in ihr

Als *Grundgesamtheit* im Sinne der Abschlußprüfung kann nicht die Gesamtheit aller vorhandenen Bestände angesehen werden. Der Gesetzgeber verlangt die Gliederung des Ausweises der Bestände nach Arten, also Gebäuden, Maschinen, Roh-, Hilfs- und Betriebsstoffen, Halbfabrikaten, Fertigfabrikaten, Wertpapieren, Schecks usw. Daher müssen die Bestände nach den Bilanzpositionen getrennt erfaßt und bewertet werden. Dementsprechend bilden in der Regel die Bestandsposten jeder einzelnen Bilanzpositionen eine selbständige Grundgesamtheit.

Bei der Prüfung einer Bilanzposition sind die hierzu gehörenden Bestandsposten als *Untersuchungseinheiten* und der gesetzliche Bilanzwert der Bestandsposten als *Untersuchungsmerkmal* zu definieren.

Dieses Untersuchungsmerkmal setzt sich jedoch vielfach aus zwei oder sogar mehreren Teilmerkmalen zusammen, z. B. bei Vorräten, Wertpapieren, Devisen u. a. aus der Bestandsmenge, gegebenenfalls auch Qualität (z. B. Feingehalt von Erzen)¹⁾ und dem Wert je Maßeinheit. Menge, Feingehalt, Einzelwert usw. mögen als *Prüfungsmerkmale* bezeichnet werden. Die Feststellungen über die einzelnen Prüfungsmerkmale können unabhängig voneinander vorgenommen werden, jedoch erlaubt erst das gemeinsame Ergebnis aller Prüfungen eine Aussage über das Untersuchungsmerkmal, den Bilanzwert des Bestandspostens.

Für die Genauigkeit, mit der der *Umfang der Grundgesamtheit* bei der Planung der Stichprobe berücksichtigt werden muß, gelten die gleichen Regeln, wie sie bereits oben dargelegt worden sind (s. S. 50f.).

¹⁾ *Trueblood/Cyert*, a. a. O., S. 148 ff., Prüfung eines Mengeninventars von Schüttgütern.

Es gehört weiter zur sorgfältigen Planung der Stichprobe, die in der Grundgesamtheit gegebene *Varianz* der Bilanzwerte bzw. ihrer einzelnen Prüfungsmerkmale aufgrund entweder von Anhaltspunkten aus früheren Untersuchungen oder von besonderen Vorerhebungen (s. S. 51f.) zu bestimmen und sowohl im Auswahlplan als auch bei der Berechnung des erforderlichen Stichprobenumfangs zu berücksichtigen. Bezüglich der Einzelheiten kann wiederum auf die entsprechenden Ausführungen über die Bestimmungsfaktoren der Schätzungen von Anteilwerten verwiesen werden.

Bei der Ermittlung der Varianz ergeben sich oft praktische Schwierigkeiten. Es kann nicht geleugnet werden, daß die Berechnung dieses Wertes insbesondere bei umfangreichen und nach komplizierten Auswahlverfahren zusammengestellten Stichproben eine etwas mühsame Arbeit ist, denn es müssen Schätzwerte der Varianzen in und zwischen allen gebildeten Auswahlgesamtheiten berechnet werden. Hierfür sind daher eine Reihe von Kurzverfahren entwickelt worden; hierzu sei auf die statistischen Lehrbücher verwiesen.

Auswahlgrundlagen und Auswahlverfahren

Die Bestände sind ihrer Natur nach sehr vielfältig; sie umfassen sowohl Grund- und Rohstoffe wie hochkomplizierte Fertigfabrikate, Kassen- und Devisenbestände sowie Wechsel-, Scheck- und sonstige Wertpapierbestände. Entsprechend vielgestaltig sind auch die *Auswahlgrundlagen*. Hierfür kommen u. a. Lagepläne mit Inventurfeldern, Lagerregale, Fässer, Tanks und andere Behälter, ferner Listen (Inventar-, Salden-, Effektenlisten etc.) und Karteien (z. B. Lagerkartei) in Frage.

Vielfältig sind auch die möglichen *Auswahlverfahren*. Allgemein wurden diese Verfahren bereits früher behandelt (s. S. 52).

Auch hier soll nicht näher auf die von der Theorie entwickelten Auswahlgrundsätze eingegangen werden, vielmehr wird am Schluß dieses Kapitels ein praktisches Beispiel für die schrittweise Anpassung eines Stichprobenverfahrens aufgrund jeweils verbesserter Informationen gegeben (s. S. 69ff.).

Sicherheitsgrad und Vertrauensbereich

Hinsichtlich der Festlegung des prüferisch notwendigen *Sicherheitsgrades* kann auf die Ausführungen über die Bedeutung der Ergebnisse der Prüfung der Sicherheit der Rechnungslegung für den Umfang der materiellen Prüfungen verwiesen werden (s. S. 53ff.). In den *Beispielen* der Literatur werden — allerdings ohne von subjektiven Momenten freie Begründungen — in der Regel Sicherheitsgrade von 90% (d. h. $t = 1,64$) oder 95% (d. h. $t = 1,96$), manchmal auch von 99,7% (d. h. $t = 3$) für erforderlich gehalten¹⁾.

Auch für die Festsetzung des prüferisch zulässigen *Vertrauensbereichs* begegnet man in der Literatur keinem objektiven Maßstab, sondern findet statt dessen regelmäßig den Hinweis auf das fachliche Ermessen des Abschlußprüfers. In den veröffentlichten *Beispielen* wird der subjektiv für zulässig erachtete Vertrauensbereich in der Regel mit rd. 1% bis 3% des Gesamtwertes der Grundgesamtheit beziffert²⁾.

¹⁾ So z. B. bei Vance/Neter, a. a. O., S. 256; Trueblood/Cyert, a. a. O., S. 88, 111, 163; Brown, a. a. O., S. 51.

²⁾ S. Fußnote 1.

Es erscheint jedoch möglich, den Vertrauensbereich für die materiellen Prüfungen dem Einzelfall angemessener und überdies objektiver festzusetzen, wenn man den Zweck der Abschlußprüfung — nämlich *Aktionärs- und Gläubigerschutz* — in Betracht zieht. Da diese Überlegungen aber für alle materiellen Prüfungen einschließlich derjenigen der Geschäftsvorfälle gelten, soll auf sie erst abschließend näher eingegangen werden (s. S. 87 ff.).

d) Beispiele

Anschließend sollen einige Beispiele aus der Literatur für die praktische Durchführung stichprobenweiser Inventurerhebungen wiedergegeben werden, und zwar für

1. Vollaufnahme und stichprobenweise Bewertung der Bestände
2. Stichprobenweise Aufnahme und Bewertung der Bestände
3. Stichprobenweise Prüfung einer permanenten Inventur
4. Schrittweise Anpassung eines Stichprobenverfahrens an die Struktur der Grundgesamtheit aufgrund schrittweise verbesserter Informationen

Beispiel 1

Vollaufnahme und stichprobenweise Bewertung der Bestände

Eine stichprobenweise Inventur haben *Bujack/Roth* in einem Artikel über „Schätzverfahren in der Inventur“¹⁾ für den Fall der Vollaufnahme und der stichprobenweisen Bewertung (Stichprobenaufbereitung) beschrieben.

Sie befassen sich in ihrem Beispiel speziell mit den Beständen im Werkstättenbereich, weil sie festgestellt hatten, daß dieser Teil der Inventurbereiche die meiste Zeit der Kalkulatoren beanspruchte und die größten Kosten verursachte. Demgegenüber würde, so meinen die Verfasser, die Inventur der Lagerbestände mittels Stichproben kaum zu wirtschaftlich interessanten Ergebnissen führen, wenn man die Möglichkeiten der permanenten Inventur sowie das Lochkartensystem voll ausnutzte²⁾.

Sie untersuchten zunächst die Struktur der Inventurkollektive in den Aufnahmebereichen und kamen für die Bereiche „Werkstätten“ und „Vor- und nachgeordnete Dienststellen“ zu folgendem Ergebnis:

„Schon eine oberflächliche Beschäftigung mit der zu untersuchenden Masse (Kollektiv) lehrt, daß auch bei den Inventurbeständen dieser Bereiche das bekannte Lagerphänomen auftritt, daß

die überwiegende Mehrzahl der Einzelpositionen einen verschwindend geringen Wertanteil repräsentiert,

während umgekehrt eine kleine Anzahl von Posten den überwiegenden Wertanteil verkörpert

und daß die Anzahl der Posten mit dem Anwachsen des Wertes des einzelnen Postens kontinuierlich abnimmt.“

¹⁾ Der Betrieb, Jahrgang 12 (1959), Nr. 21, S. 574 ff., und Nr. 22, S. 601 ff.; s. ferner die gleichen, Mathematisch-statistische Methoden bei der Inventur der Werkstattbestände, hrsg. v. d. Deutschen Arbeitsgemeinschaft für statistische Qualitätskontrolle beim AWF, Frankfurt/M., 1962.

²⁾ a. a. O., S. 577; vgl. hierzu aber das unten, S. 67f., besprochene Beispiel eines Nachweises von Inventarwerten durch Stichproben aus Aufzeichnungen der permanenten Inventur und die sinngemäß auch hierauf anzuwendenden Ausführungen über die Prüfung von Buchführungsergebnissen bei Verwendung selbsttätig arbeitender Geräte, S. 82ff.

Bujack/Roth untersuchten drei Methoden, um die ursprünglich positionsweise Aufnahme und Einzelbewertung aller Bestandsposten zu ersetzen. Die Untersuchung erfolgte nach drei Methoden, einem „Wertstaffelverfahren“, einem „Verfahren der statistischen Relation“ und dem „Verfahren der geschichteten Stichproben“.

Bei dem *Wertstaffelverfahren* wurden die geringwertigen Bestandsposten nach Wertklassen eingruppiert und die Zahl der Posten der Wertklassen mit einem Mittelwert der Wertklasse multipliziert. Bei dem Verfahren der *statistischen Relation* ging man von der empirisch festgestellten Tatsache aus, daß die geringwertigen Bestandsposten einen konstanten Anteil der Gesamtbestände ausmachen. Hier wurden nur die höherwertigen Bestandsposten aufgenommen und mit Hilfe der vorgenannten Relation auf den Gesamtwert geschlossen, ohne daß auch nur eine stichprobenweise Aufnahme der geringwertigen Posten stattfand.

Bei der *geschichteten Stichprobe* wurden die geringwertigen Posten (Inventurposten mit einem Bilanzwert bis DM 100,—) in den Werkstätten an Ort und Stelle von einem Kalkulator durch Schätzung ermittelt, gekennzeichnet, laufend numeriert und ihre Zahl festgestellt. Damit wurde die Auswahlgrundlage für die aus dieser „Schicht 1“ zu ziehende Stichprobe erstellt. Alle höherwertigen Posten wurden normal aufgenommen und die Aufnahmeergebnisse in Inventurbelegen festgehalten. Sie sollten jedoch nur stichprobenweise bewertet werden. Dafür war die Bildung von vier weiteren Schichten vorgesehen.

Die Zuordnung jedes höherwertigen Postens zu einer dieser Schichten erfolgte im Kalkulationsbüro an Hand der Inventurbelege. Aus den so geordneten Inventurbelegen für die höherwertigen Bestandsposten wurden die Stichprobeneinheiten in den einzelnen Schichten nach Maßgabe des Stichprobenplans ausgewählt.

Das *Ergebnis* einer solchen Stichprobe ist nachstehend wiedergegeben (Tabelle 9). Zur Kontrolle hatte eine exakte, d. h. nach herkömmlichen Methoden durchgeführte Inventarbewertung stattgefunden, deren Ergebnisse ebenfalls in der nachstehenden Tabelle (Spalte 2) enthalten sind, so daß ein Vergleich beider Verfahren möglich ist.

Tabelle 9: Zusammenfassung der Untersuchungen über die „Geschichtete Stichprobe“ (Bereich Werkstätten)*) 1)

	Anzahl der Posi- tionen	Exakter Inventur- wert $N\mu$ DM	Errechneter Inventur- wert $N\bar{x}$ DM	Diff. zwischen exaktem und errechn. Inven- turwert (Sp. 3 ·/. Sp. 2) in % DM		$(\bar{x} + 2\sigma_x)$ × N oberer Grenz- wert DM	$(\bar{x} - 2\sigma_x)$ × N unterer Grenz- wert DM	Abweichung vom exakt. Inventur- wert i. % Sp. 6 u. 7 bezogen auf 2 %
	1	2	3	4	5	6	7	8
Kostenträger 1	18 568	4 194	4 189	—0,117	—4,9	4 203	4 176	+ 0,200/—0,434
Kostenträger 2	5 759	860	859	—0,109	—0,9	868	850	+ 0,890/—1,108
Kostenträger 3	8 298	2 059	2 059	—0,017	—0,4	2 068	2 050	+ 0,437/—0,472
Kostenträger 4	9 895	1 743	1 741	—0,154	—2,7	1 751	1 731	+ 0,407/—0,714
Kostenträger 5	9 491	2 175	2 176	+ 0,019	+ 0,4	2 187	2 165	+ 0,525/—0,492
	52 011	11 032	11 024	—0,077	—8,4	11 076	10 972	+ 0,396/—0,549

*) Geldwertangaben der Spalten 2, 3, 5, 6, 7 in DM 1 000,—; Abweichungen der Fußzeile gegenüber den Spaltensummen beruhen auf Rundungen der Zahlen.

$\bar{x}$ = Stichprobenmittelwert für das Gesamtkollektiv

1) *Bujack/Roth* (1959), S. 603.

Über die mit den erwähnten Verfahren erzielten *Rationalisierungserfolge* berichten *Bujack/Roth*, allerdings ohne sich über etwaige Unterschiede in der Genauigkeit und Sicherheit der Ergebnisse zu äußern, wie folgt:

„Zusammengefaßt sollen bei vorsichtigem Ansatz die Erfolge der einzelnen Verfahren gegeneinander abgewogen werden. Sie beziehen sich auf etwa eine Viertel-million Inventurposten bei einer Totalaufnahme.

Methode des Wertstaffelverfahrens: Der Zeitgewinn liegt bei 8 Tagen, die Kostenersparnis dürfte die Grenze von DM 100 000,— nicht überschreiten.

Methode der statistischen Relation: Der Zeitgewinn liegt bei rd. 14 Tagen, die Kostenersparnis bei rd. DM 150 000,—. Diese Vorteile werden aber z. B. mindestens alle 5 Jahre ganz oder teilweise aufgehoben (weil dann eine Vollerhebung notwendig wird, um festzustellen, ob der Anteil der geringwertigen Bestandsposten am Gesamtbestand sich geändert hat, Anm. d. Verf.).

Methode der geschichteten Stichprobe: Der Zeitgewinn liegt durch frühere Beendigung der Bewertungsarbeiten bei 3 Wochen und hat eine Kostensenkung von etwa DM 200 000,— zur Folge.“

Beispiel 2

Stichprobenweise Aufnahme und Bewertung der Bestände

Selbst dort, wo eine als Basis für eine Auswahlgrundlage in Frage kommende Ordnung der Einheiten nicht von vornherein gegeben ist, läßt sie sich häufig mit verhältnismäßig geringem Aufwand herstellen, wie das folgende Beispiel der „Bewertung des Inventars körperlicher Gegenstände“ zeigt, das *Deming*¹⁾ ausführlich beschreibt. Er zeigt die Möglichkeit stichprobenweiser Bestandsaufnahme und -bewertung an einem besonders schwierigen Fall, nämlich einer Inventur der in der Produktion befindlichen Gegenstände (Halbfabrikate). Über die Arbeiten zur Aufstellung der *Auswahlgrundlage* läßt sich dem Bericht folgendes entnehmen:

Die Inventur des Fertigungsmaterials hatte in vier Gebäuden mit 15 Abteilungen zu erfolgen. Rohmaterial (Gummi, Bleche, Draht, Silber, Gold), unfertige und fertige Teilstücke durchliefen dort eine Fabrikationsstufe nach der anderen. Auf jeder folgenden Fabrikationsstufe war das Teil, Material oder Montagestück wertvoller als auf der Vorstufe.

Die Teile, Materialien und Montagestücke wurden an einem Wochenende nach Fertigungsstillstand in „Partien“ aufgeteilt. Als Partie galt eine Gruppe identischer Teile, die in einem oder mehreren Behältern (Pfanne, Mulde, Korb, Fach) enthalten waren. Um eine aus mehreren Mulden, Fächern u. ä. bestehende Partie abzugrenzen, benutzte man weiße Klebestreifen. Es gab ungefähr 43 400 Partien von Materialien verschiedenster Art, deren Wert zwischen ungefähr \$ 10 und \$ 9000 lag.

Die Aufteilung der Bestände in Partien erfolgte in zwei Schritten. Zunächst mußte jeder Werkmeister einer Abteilung einige Stunden vor Beginn der Inventur erscheinen, um die Partien zu bilden und zu bestimmen, die nach seiner Meinung einen Wert von \$ 500 oder mehr hatten (Schicht 1). Er hatte diesen Partien rote Zettel anzuheften und sie fortlaufend zu numerieren. Das mußte geschehen sein, bevor die Arbeiter erschienen, um den zweiten Schritt durchzuführen.

¹⁾ *Deming* (1960), S. 150 ff.

Je zwei Arbeiter hatten alsdann das übrige Material, also mit Werten unter \$ 500 (2. Schicht) in Partien aufzuteilen und je 20 Partien zu einer „Zone“ zu vereinigen. Die Partien in einer Zone erhielten die Seriennummern 1 bis 20. Zur Numerierung wurden Notizblöcke von je 20 Blatt, vorgedruckt mit den Seriennummern 1 bis 20, verwandt. Die 20 Blatt mußten dann auch die Nummer der Zone tragen. Jede Aufnahmegruppe (2 Arbeiter) erhielt ein Paket solcher Blöcke und hatte je einen Block für eine Zone zu verwenden. Die einmal einer Partie zugewiesenen Seriennummern mußten erhalten bleiben.

Es war jedoch zu diesem Zeitpunkt den Arbeitern immer noch erlaubt, ihr Material in gebräuchlichere Partien umzugruppieren oder den Werkmeister auf jedes Material aufmerksam zu machen, das ihrer Meinung nach eine eigene Partie von \$ 500 oder mehr bilden sollte. Nach diesem Schritt aber war dann keine Änderung der Partien oder der Ordnungsnummern mehr gestattet. Dieser Schritt schuf die Auswahlgrundlage.

Untersuchungseinheit war die Partie.

Deming beschreibt anschließend, wie aus der so geordneten Grundgesamtheit nun die Zufallsauswahl der Partien, ihre körperliche Bestandsaufnahme und durch die Kalkulatoren mit Hilfe der Kalkulationsunterlagen die Bewertung der in die Stichprobe einbezogenen Partien der zweiten Schicht (unter \$ 500) sowie aller Bestandsposten der ersten Schicht (\$ 500 und mehr) erfolgte. Eingehend werden von ihm sodann auch das Schätzverfahren und die Auswertung der Ergebnisse erläutert.

Als *Rationalisierungserfolge* seines Verfahrens nennt *Deming* u. a.: „Weniger Aufnahmepersonal erledigte die Arbeit (ungefähr $\frac{1}{5}$ des Personals bei vollständiger Inventur) . . .“ und ferner

„Die vollständige Inventur erforderte in den vergangenen Jahren zwei Wochen für die Aufnahme selbst und etwa vier Monate für die Auswertung. Der größte Teil der Auswertung geschah in Überstunden mit der dabei üblichen Zahl von Fehlern. Im Gegensatz dazu benötigte diese Stichprobeninventur einen Tag und ein paar Tage für die Auswertung.“

Ähnliche Beispiele bringen u. a. auch *Courtright/Procassini*¹⁾, *Rudell*²⁾, *Obrock*³⁾ nicht zuletzt auch *Vance/Neter*⁴⁾ und *Trueblood/Cyert*⁵⁾. Es ist jedoch nicht möglich, hierauf im einzelnen einzugehen. Eine Bibliographie der einschlägigen Literatur wird vom American Institute of Certified Public Accountants herausgegeben⁶⁾.

1) *Courtright, W. E. und Procassini, A. A.*, Inventory Valuation by Sampling, in *Industrial Quality Control*, Februar 1958, S. 16 ff.

2) *Rudell, A. L.*, Applied Sampling Doubles Inventory Accuracy, Halfs Cost, in *Bull. Nat. Ass. Accountants*, Oktober 1957, S. 5 ff.

3) *Obrock, R. F.*, A Case Study of Statistical Sampling, *JofA*, März 1958, S. 53 ff.

4) a. a. O., S. 260 ff.

5) a. a. O., S. 85 ff. 148 ff., 158 ff.

6) American Institute of Certified Public Accountants, *Glossary of Statistical Terms for Accountants and Bibliography on the Application of Statistical Methods to Accounting, Auditing and Management Control*, New York 16, N. Y., Sept. 8, 1958.

Beispiel 3

Stichprobenweise Prüfung einer permanenten Inventur

Stichprobenverfahren können nicht nur — wie in den vorstehenden Beispielen — bei Stichtagsinventuren, sondern auch zur Prüfung der permanenten Inventur benutzt werden, wie die folgenden Beispiele zeigen. Hierbei gelangt man auf retrograde Weise, d. h. ausgehend von den Inventuraufzeichnungen des Unternehmens, zu den interessierenden Feststellungen.

Zunächst sei auf das von *Trueblood/Cyert*¹⁾ wiedergegebene Beispiel für den „Nachweis von Inventarwerten durch Stichproben aus Aufzeichnungen der permanenten Inventur“ eingegangen:

In diesem Fall fand die Stichprobenerhebung in einer Handelsfirma mittlerer Größe statt. Aus innerbetrieblichen Kontrollgründen war das Inventar in zwei Teile untergliedert. Die „hochwertigen Bestände“ setzten sich aus ca. 800 Einzelposten zusammen, die schätzungsweise 50% bis 60% des gesamten Inventarwertes ausmachten. Der andere Teil des Inventars, als geringwertig bezeichnet, bestand aus ungefähr 4800 verschiedenen Einzelposten.

Aus besonderem Anlaß sollte die *Lagerbuchführung* für einen bestimmten Stichtag überprüft werden, ohne daß eine Vollerhebung der geringwertigen Bestandsposten stattfinden mußte.

Ursprünglich hatte das Unternehmen geplant, eine totale körperliche Aufnahme der hochwertigen Bestände vorzunehmen und die geringwertigen Bestände in einer bewußt ausgewählten Stichprobe körperlich so weit aufzunehmen, wie die Zeit es zuließ. Eine vollständige körperliche Aufnahme aller Bestände wäre unter den gegebenen Umständen fast unmöglich gewesen.

In dem Unternehmen war es Übung, die Ergebnisse der permanenten Inventur, mit der die Richtigkeit der Lagerbuchführung laufend überprüft wurde, monatlich in einem „Gewinn- und Verlust“-Bericht zusammenzufassen. Die Berichte zeigten den *Dollarbetrag der einzelnen Überschüsse und Fehlmengen* auf, die sich aus der körperlichen Aufnahme gegenüber dem Buchwert ergaben, und bezogen diese Dollarwerte auf die Werte der entsprechenden Inventarposten. Diese Berichte, die seit Jahren regelmäßig angefertigt wurden, wurden sowohl von dem Unternehmen wie seitens der Wirtschaftsprüfer als hinreichend zuverlässiger Maßstab für den kombinierten „Fehlerfaktor“ aus Lagerbuchführung und körperlicher Inventur angesehen. Die Erfahrung zeigte, daß es keine größeren Varianzen im festgestellten Fehlerfaktor aus saisonalen oder anderen Gründen gab.

Um die Fehler in der Lagerbuchführung zu schätzen, wurde beschlossen, die *geringwertigen Posten mittels einer Stichprobe* körperlich aufzunehmen, den *Fehlerfaktor dieser Stichprobe* zu ermitteln und diesen Fehlerfaktor mit den auf gleiche Weise berechneten Fehlerfaktoren verschiedener früherer „Gewinn- und Verlust“-Berichte zu vergleichen.

Hierbei ergab sich, daß der aufgrund der Stichprobe ermittelte Fehlerfaktor eine *Hochrechnung* des Stichprobenergebnisses, d. h. eine hinreichend genaue *Schätzung des Gesamtwertes* der 4800 geringwertigen Inventarposten, erlaubte.

¹⁾ a. a. O., S. 85 ff.

Der Gesamtwert der geringwertigen Bestände wurde berechnet, indem die Sollwerte der geringwertigen Bestände laut Lagerbuchführung um den Fehlerfaktor berichtigt wurden, der aus der Stichprobe abgeleitet worden war und der seiner Größe nach den aus den früheren „Gewinn- und Verlust“-Berichten berechneten Fehlerfaktoren entsprach.

Über die *Ergebnisse des Verfahrens* wurde folgendes mitgeteilt:

„Der Schätzwert für die Differenz zwischen körperlicher Aufnahme und den Aufzeichnungen der permanenten Inventur, den die Zufallsstichprobe der geringwertigen Posten lieferte, war ein Gewinn (d. i. mengenmäßiger Überschuß) von 0,8%. Die in dieser Berechnung enthaltene Varianz war derart, daß die Statistiker sagen konnten, die tatsächliche Größe des Gesamtfehlers liege mit 95% Sicherheit zwischen einem „Gewinn“ von 4,2% und einem „Verlust“ von 2,6%.“

Dieses Ergebnis wird wie folgt erläutert:

„... Angenommen, der Wert der geringwertigen Posten habe laut permanenter Inventur \$ 1 000 000 betragen, so hätte aufgrund der vorausgegangenen permanenten Inventur, die durch eine zufällige körperliche Prüfung bestätigt worden ist, festgestellt werden können, daß der tatsächliche Inventarwert wahrscheinlich in den Bereich von \$ 974 000 bis \$ 1 042 000 fallen würde. Diese Untersuchung gab den wahrscheinlichen Dollarbetrag der geringwertigen Inventarbestände mit bemerkenswert größerer Genauigkeit an, als sie mit den üblicheren Prüfungsverfahren hätte erzielt werden können.“

Darüber hinaus werden folgende *weitere Vorteile* des Verfahrens festgestellt:

„Die Einsparungen beliefen sich auf mehrere tausend Dollar an Löhnen, die anderenfalls für die Inventurarbeiten hätten gezahlt werden müssen. Weit wichtiger war jedoch, daß eine bedeutende Zeiteinsparung erreicht wurde. Eine vollständige Inventur der geringwertigen Bestände wäre im Rahmen der vorgegebenen Zeit fast unmöglich gewesen.“

Die 800 hochwertigen Bestandspositionen hätten ebenfalls mit Stichproben untersucht werden können. Von den Unternehmensvertretern wurde jedoch entschieden, diese im Einzelwert sehr unterschiedlichen Posten total aufzunehmen. Wäre das Unternehmen, meinen die Autoren, mit der Anwendung statistischer Verfahren vertraut gewesen, wäre wahrscheinlich selbst für die hochwertigen Posten statt der Vollerhebung eine geschichtete Stichprobe angewandt worden.

Ebenso wie die Ermittlung und Prüfung körperlicher Bestände durch Stichproben ist auch die Ermittlung und Prüfung *nicht körperlicher Bestände*, also insbesondere der verschiedenartigen Forderungen und Verbindlichkeiten, stichprobenweise möglich. Ein besonders instruktives Beispiel hat *Deming* für die „Schnellprüfung von Konten in Hauptbüchern“ veröffentlicht, auf das später eingegangen werden soll, weil Ziel dieser Erhebung nicht in erster Linie die Feststellung des Gesamtwertes eines sehr großen Debitorenbestandes, sondern vornehmlich die Ermittlung der unterschiedlichen Besetzung einzelner Wertklassen (Schichten) war. An dieser Stelle genügt die Feststellung, daß auch Bilanzwerte von nicht körperlichen Beständen zweckentsprechend mit Hilfe von Stichproben geschätzt werden können.

Beispiel 4

Schrittweise Anpassung eines Stichprobenverfahrens an die Struktur der Grundgesamtheit aufgrund schrittweise verbesserter Informationen

R. F. Obrock¹⁾ schildert die schrittweise Anpassung eines Stichprobenverfahrens aufgrund jeweils verbesserter Informationen.

Ziel war die *Berichtigung des Buchwertes des Inventars*, bestehend aus Labor-, Instandhaltungs- und Papiervorräten für den Unternehmensbedarf. Von diesem Inventar waren folgende Kennzahlen bekannt:

Zahl der Inventarposten	3 900
Gesamtzahl der Einheiten aller Inventarposten	260 000
Inventarwert	\$ 185 000
Lagerumschlag pro Jahr	2,6
Kleinster Einzelwert eines Postens	\$ 0,005
Größter Einzelwert eines Postens	\$ 120

In den vorangegangenen Jahren wurde jeder Inventarposten gezählt und die Menge mit dem Kontenausweis verglichen. Sofern Abweichungen auftraten, wurden sie geprüft und die entsprechenden Konten berichtigt. Eine Zusammenstellung der von 1951 bis 1955 erforderlichen Wertberichtigungen des Inventars ergab folgendes Bild:

Tabelle 10: Entwicklung der Netto-Wertberichtigungen der Inventarwerte in der Zeit von 1951 bis 1955²⁾

Jahr	Netto- Wertberichtigung in \$	Posten mit	
		Überschüssen in %	Fehlmengen in %
1951	./.. 260	12	20
1952	./.. 406	9	13
1953	./.. 1 026	9	19
1954	./.. 1 339	11	22
1955	./.. 566	11	21

Für die *Stichprobenplanung* waren folgende Überlegungen maßgebend: Eine besonders schwundempfindliche Gruppe von Posten sollte lückenlos geprüft werden. Diese Gruppe umfaßte 27 Inventarposten. Weiter sollte der Stichprobenplan nach statistischen Grundsätzen mit dem kleinstmöglichen Stichprobenumfang zu befriedigender Genauigkeit des Ergebnisses führen und trotzdem sowohl leicht zu verwirklichen sein als auch den Prüfungsgrundsätzen Rechnung tragen. Schließlich hielt man es für erforderlich, bestimmte Teilgebiete des Inventars stärker als für die gewünschte Genauigkeit notwendig zu erfassen, um Unterlagen für künftige Erhebungen zu erhalten. Es hatte sich nämlich herausgestellt, daß die meisten Merkmale der Grundgesamtheit unbekannt waren.

Zur Schätzung der „*Nettowertberichtigung*“ der Inventarwerte an Hand der Stichprobe wurde die Wertberichtigung der Inventarposten zum Untersuchungsmerkmal gemacht. Stimmte die festgestellte Menge mit dem Kontenausweis

¹⁾ A Case Study of Statistical Sampling, JofA, März 1958, S. 53 ff.

²⁾ Die Tabellen 10 bis 15 sind dem Aufsatz von Obrock entnommen.

überein, so betrug die Wertberichtigung Null. Ein festgestellter Mengenüberschuß gegenüber dem Kontenausweis führte zur Berichtigung in Höhe des Produktes aus Mengenüberschuß und mittlerem Selbstkostenbetrag mit positivem Vorzeichen. Fehlmengen wurden ausgebucht, ergaben also negative Wertberichtigungen.

Zu der Methode, nur die Nettowertberichtigung zu schätzen, nicht aber den berichtigten Gesamtwert zu ermitteln, muß festgestellt werden, daß der Abschlußprüfer die Nettowertberichtigung nur dort schätzen darf, wo prüferisch sichergestellt ist, daß die ausgewiesenen Werte der einzelnen Inventarposten richtig addiert wurden. Anderenfalls könnte ein einzelner erheblicher Fehler sowohl vom Buchführungssystem als auch von der Stichprobe unentdeckt bleiben¹⁾).

Obwohl die Aufbereitung der Daten über die Wertberichtigungen von 1955 für die Stichprobenplanung unzureichend war, ließen sich mit Hilfe der in die Kontenkarten (Lochkarten) eingelochten Merkmale allgemeine Folgerungen für eine *Schichtenbildung* nach der Größe der Selbstkosten ziehen. Die Gruppierung der Inventarposten nach ihren Selbstkosten und die Prüfung der Wertberichtigungen in den verschiedenen Selbstkostengruppen führten zu der Erkenntnis, daß die durchschnittliche Wertberichtigung unabhängig vom Selbstkostenbetrag war und nahe Null lag und daß ferner die Varianz der Wertberichtigungen um den Mittelwert mit wachsender Selbstkostengröße zunahm.

Aufgrund dessen wurden die in Tabelle 11 beschriebenen Schichten gebildet:

Tabelle 11: Schichtungsplan (1. Verfahren)

Schicht	Schichtenmerkmal bzw. Schichtengrenze *)	Schichten- umfang
1	Kein Wert oder Selbstkosten = 0	73
2	Über Null bis unter \$ 1	2 077
3	\$ 1 bis \$ 4,99	1 186
4	\$ 5 bis \$ 19,99	432
5	\$ 20 und darüber	99
6	Verlustempfindliche Einheiten	27
Insgesamt		3 894

*) Die Schichten 1 bis 5 schließen alle diejenigen Einheiten der entsprechenden Wertklasse aus, die als empfindliche Einheiten vorher ausgewählt und in Schicht 6 zusammengefaßt worden waren.

Für diese *Wahl der Schichten* werden folgende Gründe angeführt: Zunächst habe man drei Schichten für ausreichend gehalten, um die meisten der durch eine Schichtung zu erzielenden Vorteile zu erlangen. Da es sich aber — wie gesagt — um grundlegende Arbeiten für künftige Stichprobenerhebungen gehandelt habe, habe etwas mehr über die Merkmale der Grundgesamtheit in Erfahrung gebracht werden sollen. Aus diesem Grunde seien feinere Unterscheidungen vorgenommen worden.

„Schicht 1 ist ein Beispiel für die Abgrenzung von Teilen, über die zusätzliche statistische Informationen benötigt wurden“, schreibt *Obrock* dann und fährt fort: „In der Grundgesamtheit gab es 73 Posten, deren Lochkarten entweder den Selbstkostenbetrag Null auswiesen oder keine Selbstkostenangaben enthielten.

¹⁾ So auch *Mackenzie*, a. a. O., S. 525.

In unserem Buchführungssystem konnte dieser Fall entweder als Folge der Maschinenberechnung eines Nullsaldos oder der Eröffnung eines neuen Kontos für einen noch nicht auf Lager befindlichen Posten eintreten. Man glaubte, daß eine Debet- oder Kreditberichtigung unter diesen Posten im Grunde nicht vorkommen könne (trotz der Tatsache, daß jeder dieser Posten normalerweise einen Selbstkostenbetrag aufweist, der hoch oder niedrig sein kann). Wir hielten es für wichtig, diese Annahme durch Bildung einer eigenen Schicht der „Bestandslosen Posten“ („zero or no unit cost“ items) zu prüfen. Ein ähnlicher Wunsch nach Verfeinerung im Rahmen dieser experimentellen Stichprobenerhebung veranlaßte uns, den Rest des Selbstkostenbereichs in vier weitere Schichten zu teilen. Schließlich ergab die Forderung, die schwundempfindlichen Posten lückenlos zu erheben, die sechste Schicht.“

Es zeigte sich, daß die Wertberichtigungen mit wachsenden Selbstkosten weiter streuten, so daß die optimale Aufteilung sich als die geeignete Methode herausstellte, wobei der Auswahlatz im Maß der Zunahme der Selbstkosten ausgedehnt wird.

Aufgrund der bisherigen Erkenntnisse wurde der in Tabelle 12 wiedergegebene Schichtungs- und Auswahlplan beschlossen:

Tabelle 12: Schichtungs- und Auswahlplan (1. Verfahren)

Schicht	Umfang der			Ordnungsnummern in der Grundgesamtheit
	Schichten	Stichprobe		
		absolut	in % von N	
1	73	40	55	0 000 bis 0 072
2	2 077	175	8	0 073 bis 2 149
3	1 186	225	19	2 150 bis 3 335
4	432	150	35	3 336 bis 3 767
5	99	99	100	nicht benötigt
6	27	27	100	nicht benötigt
Insgesamt	3 894	716	18	

Obwohl damit die Planung der aktuellen Stichprobe abgeschlossen war, ist es interessant zu verfolgen, wie die Ergebnisse dieser Stichprobe dazu benutzt wurden, einen *noch wirkungsvolleren Plan* für die zukünftigen Erhebungen zu entwerfen.

Die Ergebnisse der Stichprobenprüfung ließen sich wie folgt zusammenfassen:

Tabelle 13: Stichprobenergebnisse (1. Verfahren)

Schicht	Durchschnittliche Wertberichtigung je Stichprobeneinheit in der Schicht	Zahl der Untersuchungseinheiten je Schicht	Geschätzte Wertberichtigung je Schicht und insgesamt	Standardfehler je Schicht
1	\$ ± 0,000	73	\$ ± 0	0,00
2	— 0,214	2 077	— 445	3,63
3	— 0,609	1 186	— 722	4,60
4	+ 0,470	432	+ 203	6,76
5	— 0,779	99	— 77	17,86
6	— 2,367	27	— 64	12,66
Insgesamt	—	—	1 105	—

Diese Ergebnisse werden von *Obrock* wie folgt interpretiert: „Die Zahlen der durchschnittlichen Wertberichtigungen . . . bestärkten uns in der Meinung, daß zwischen ihnen und den Selbstkosten keine enge Beziehung bestand. Mit Ausnahme von Schicht 6 ist die Streuung der durchschnittlichen Wertberichtigungen von Schicht zu Schicht klein. Schicht 6 hat eine erwartungsgemäß große durchschnittliche Wertberichtigung, da die Posten dieser Schicht nach Maßgabe der Schwundwahrscheinlichkeit vorab ausgewählt worden waren.“ Und im Hinblick auf die Standardabweichungen in den Schichten schreibt er u. a.: „Wir stellten zunächst fest, daß die Standardabweichung von Schicht 4 auf Schicht 5 stark anstieg. Die Zunahme von Schicht 2 bis Schicht 4 war gemäßigter und brachte uns auf den Gedanken, daß wir einige dieser Schichten ohne großen Genauigkeitsverlust zusammenlegen könnten. Auch der große Unterschied zwischen Schicht 1 und Schicht 2 war angesichts der verhältnismäßig kleinen Zahl der Einheiten der Schicht 1 im Vergleich zum Umfang der Schicht 2 unwesentlich. Wir beschlossen daher, Schicht 5 unverändert bestehenzulassen, die Schichten 3 und 4 sowie die Schichten 1 und 2 zu je einer neuen Schicht zusammenzufassen.“

Damit ergaben sich die in Tabelle 14 aufgeführten *neuen Schichtengrenzen und -umfänge*.

Tabelle 14: Schichtungsplan (2. Verfahren)

Neue Schicht	Schichtenmerkmal bzw. Schichtengrenzen	Schichtenumfang
1	Unter \$ 1	2 150
2	\$ 1 bis \$ 19,99	1 618
3	\$ 20 und mehr	99
4	Schwundempfindliche Einheiten	27
Insgesamt		3 894

Bei optimaler Aufteilung, die sich als zweckmäßig erwiesen hatte, war die „neue Schicht 2 . . . mit einem Auswahlsatz von 20% in der Stichprobe zu berücksichtigen, gegenüber einem Auswahlsatz von 13% für die neue Schicht 1, was ungefähr dem Verhältnis von \$ 5,28 : \$ 3,57 der Standardabweichungen in den beiden Schichten entspricht. Wären wir diesen Verhältnissen streng gefolgt, hätten wir die neue Schicht 3 mit einem Auswahlsatz von ungefähr 65% berücksichtigt. Ist aber einmal ein so hoher Auswahlsatz bei einer Schicht kleinen Umfangs erreicht, so ist es oft leichter, sie lückenlos zu prüfen. Die Schicht der schwundempfindlichen Einheiten würde natürlich immer lückenlos geprüft. Da die beiden neuen Schichten 3 und 4 lückenlos zu prüfen sind, könnte man sie in einer Schicht zusammenfassen. Wir hielten es aber für wertvoll, die Feststellung der einzelnen Mittelwerte und Standardabweichungen für diese Gruppen als Hilfe für die Planung zukünftiger Stichprobenerhebungen beizubehalten.“ Es ergab sich daher der in Tabelle 15 beschriebene *neue Stichprobenplan*.

Tabelle 15: Auswahlplan aufgrund verbesserter Informationen

Neue Schicht	Standard- fehler in \$	Umfang der		
		Schichten	Stichprobe	
			absolut	in % des Schichtenumfangs
1	3,57	2 150	275	13
2	5,28	1 618	325	20
3	17,86	99	99	100
4	12,74	27	27	100
Insgesamt		3 894	726	19

Obrock stellt schließlich noch einen *Vergleich* zwischen der Zahl der Schichten, der Größe des Vertrauensbereichs und dem erforderlichen Stichprobenumfang bei gleichem Sicherheitsgrad von 95% in allen Varianten an. Das Ergebnis kann wie folgt dargestellt werden:

Tabelle 16: Einfluß der Zahl der Schichten auf den Umfang der Stichprobe und die Ergebnissenauigkeit ¹⁾

Zahl der Schichten	Stichprobenumfang	Vertrauensbereich in \$
6	716	$\bar{X}^2) \pm 1\,300$
4	726	$\bar{X} \pm 1\,185$
2	726	$\bar{X} \pm 1\,345$
2	882	$\bar{X} \pm 1\,185$

Hieraus ergibt sich, daß bei den von ihm gebildeten vier Schichten eine Genauigkeit von $\pm \$ 1\,185$ bei einem Stichprobenumfang von 726 Einheiten erreicht werden kann (Zeile 2). Eine Zusammenfassung der Schichten der verlustunempfindlichen Einheiten und damit die Verringerung der Zahl der Schichten von vier auf zwei führt entweder zu einer Vergrößerung des Vertrauensbereichs um \$ 160 (Zeile 3 gegenüber Zeile 2) oder zu einer Steigerung des erforderlichen Stichprobenumfangs um 156 Einheiten (Zeile 4 gegenüber Zeile 2). Gegenüber der gewählten Schichtung sind daher alle anderen ungünstiger hinsichtlich der Vorarbeiten, des Stichprobenumfangs oder der Ergebnissenauigkeit.

Wie in der übrigen bisherigen Literatur, so sind auch in dem vorstehend wiedergegebenen Beispiel nicht die Überlegungen angeführt, die zur Festsetzung des Sicherheitsgrades und des Vertrauensbereiches geführt haben.

B) Bestandsanalysen

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Bei der Prüfung der Bestände kommt es unter Umständen nicht nur auf die Bestimmung von Gesamtwerten an, sondern auch auf die Aufgliederung von Beständen nach bestimmten Merkmalen, z. B. nach Größenordnungen, nach dem Grad der Realisierbarkeit bzw. der Dauer der geldmäßigen Gebundenheit usw.

¹⁾ Erstellt aufgrund der Angaben von Obrock, a. a. O., S. 58.

²⁾ $\bar{X}$ = Schätzwert für den Gesamtbetrag der Nettowertberichtigung.

Diese Analysen können ebenfalls mit Hilfe von *Stichprobenverfahren*, und zwar je nach der Art der Fragestellung mit Stichproben für *Anteilsschätzungen* oder solchen für *Mittelwertbestimmungen*, rationell durchgeführt werden.

b) Schätzung von Anteilen und Mittelwerten mittels Schätzstichproben

(1) Formelapparat

Der Formelapparat, mit dem sich die benötigten Stichprobenumfänge im Fall von *Anteilsschätzungen* ermitteln lassen, ist bereits auf Seite 40 ff. besprochen worden. Sollen *Mittelwerte* geschätzt werden, so findet grundsätzlich der auf Seite 58 ff. erörterte Formelapparat Anwendung. Allerdings ist hier der Wert e der Meßwert für den Vertrauensbereich für den Mittelwert (und nicht mehr für den Gesamtwert). Das hat zur Folge, daß der Faktor N für den Umfang der Grundgesamtheit lediglich in den Formeln auftritt, bei denen der Endlichkeitsfaktor berücksichtigt ist.

Als *Planungsformeln* haben daher bei der *Mittelwertschätzung* aufgrund einer *uneingeschränkten Zufallsauswahl* zu gelten

bei Vernachlässigung des Endlichkeitsfaktors

$$(39) \quad n \geq \frac{t^2 \sigma^2}{e^2}$$

bei Berücksichtigung des Endlichkeitsfaktors

$$(40) \quad n \geq \frac{t^2 \sigma^2 N}{e^2 (N - 1) + t^2 \sigma^2}$$

(2) Wirkung der Bestimmungsfaktoren auf den Stichprobenumfang

Infolge der Gleichartigkeit der Beziehungen zwischen den einzelnen Bestimmungsfaktoren in den Formeln für die Anteilsschätzungen hier wie im Fall der Schätzung von Fehleranteilen sind auch die Wirkungen der Faktoren auf den Stichprobenumfang dieselben, wie sie bereits auf Seite 44 ff. beschrieben worden sind. Analog kann hinsichtlich der Wirkung der Faktoren in den Formeln für die Mittelwertschätzung grundsätzlich auf die Ausführungen über die Wirkung der Faktoren in den Formeln für die Gesamtwertschätzung (s. S. 59 f.) verwiesen werden.

c) Anwendbarkeit auf das Prüfungsgebiet

(1) Prüferische Auswertung

Da die Bestimmungsfaktoren für den Stichprobenumfang sowie ihre Beziehungen zueinander, wie vorstehend festgestellt, grundsätzlich die gleichen sind, ist auch die Anwendbarkeit von Schätzstichproben hier grundsätzlich im gleichen Umfang gegeben, wie auf Seite 49 ff. bzw. Seite 60 ff. dargestellt. Auch hier sind also bei gegebenem Sicherheitsgrad und Vertrauensbereich Stichproben um so rentabler, je größer der Umfang der Grundgesamtheit und je kleiner die Varianz sind.

Während aber bei den *Anteilsschätzungen* zur Prüfung der Wirksamkeit der unternehmensinternen Kontrollvorschriften sehr kleine Fehleranteile oder solche nahe der 50%-Marke, die große und größere Stichprobenumfänge erfordern,

prüferisch von geringerer Bedeutung sind, interessieren bei der *Bestandsanalyse* die Schätzwerte der Anteile der einzelnen Merkmalkategorien grundsätzlich unabhängig von ihrer Größe, d. h. die Stichproben müssen so geplant werden, daß sowohl die kleineren als auch die größeren Anteile mit der verlangten Genauigkeit geschätzt werden können. Sind die Unterschiede zwischen den zu schätzenden Anteilen bzw. ihren Varianzen sehr groß und befinden sich auch sehr kleine Anteile unter ihnen, so ist aufgrund der auf Seite 46 f. erläuterten Zusammenhänge ein sehr großer Stichprobenumfang erforderlich. Das bedeutet aber unter Umständen gleichzeitig, daß der Umfang der Stichproben, in denen diese Anteile festgestellt werden, größer ist als bei der geforderten Genauigkeit der Schätzwerte für diese Anteile erforderlich und hierdurch die Ergebnisse für diese Anteile die geforderte Genauigkeit überschreiten.

Will der Prüfer also in einzelnen Stichprobenverfahren Anteile sehr unterschiedlicher Größe einer Grundgesamtheit ermitteln, so steht er vor dem Problem, sich zwischen der gewünschten Genauigkeit der „Grenzanteile“ einerseits und der Größe des Stichprobenumfangs andererseits entscheiden zu müssen. Hier wird es darauf ankommen, einen prüferisch tragbaren Kompromiß zwischen der erforderlichen Genauigkeit der am stärksten interessierenden Anteile und einem rationellen Umfang der Stichprobe zu finden.

Vor der gleichen Frage steht der Prüfer, wenn er in einem Verfahren die *Mittelwerte verschiedener Merkmale* schätzen will, z. B. die durchschnittliche Höhe und die mittlere Laufzeit der Forderungen. Hier muß er sich also bei unterschiedlichen Varianzen der Werte der beiden Merkmale in der Grundgesamtheit entscheiden, ob er ein ungenaueres Ergebnis bei einem Merkmal im Interesse eines geringeren Stichprobenumfangs oder umgekehrt einen unrentabel großen Stichprobenumfang in bezug auf die weniger scharfen Genauigkeitsforderungen des anderen Merkmals in Kauf nehmen will. Die Entscheidung muß er aus der Zielsetzung seiner Untersuchung treffen. Beispiele für Stichprobenerhebungen, bei denen diese Fragen eine Rolle spielen, werden auf Seite 76 ff. wiedergegeben.

(2) Grundlagen für die Festsetzung der Bestimmungsfaktoren

Als *Grundgesamtheit* kommen alle jene *Teilmengen* in Betracht, für die auch die Gesamtwerte zu ermitteln sind (s. S. 61). Die Art der in Frage kommenden *Untersuchungsmerkmale* richtet sich nach den mit der Bestandsanalyse im Einzelfall verfolgten Zielen (z. B. Ermittlung der Anteile bestimmter Arten oder Größenklassen von Vorräten am Gesamtbestand, mittlere Laufzeit der Kredite).

Der *Umfang* der Grundgesamtheit und die *Anteile* bzw. die *Varianz* in ihr oder die Anteile bzw. Varianzen in den Teilgesamtheiten sind nach den auf Seite 50 ff. resp. Seite 61 ff. dargestellten Grundsätzen zu ermitteln.

Auch hinsichtlich der *Auswahlgrundlagen* und *Auswahlverfahren* kann auf die früheren Ausführungen verwiesen werden (s. S. 51 f. bzw. 62).

Das gleiche gilt für die Festsetzung des *Sicherheitsgrades* und des *Vertrauensbereichs*, der so zu bestimmen ist, daß die Zwecke, die die Analyse im Rahmen der Abschlußprüfung zu erfüllen hat, erreicht werden.

d) Beispiele

Verfahren der Analyse von Beständen sind in der Literatur bereits mehrfach geschildert worden. Im folgenden soll eine Reihe interessanter Einzelheiten zweier Beispiele wiedergegeben werden, und zwar

1. Analyse eines Debitorenbestandes nach Umfang und Wert bestimmter Größenklassen der Debitoren und Ermittlung ihres Gesamtwerts,
2. Kreditfristanalyse in Verbindung mit einer Gesamtwertschätzung der Debitoren.

Beispiel 1

Analyse eines Debitorenbestandes nach Umfang und Wert bestimmter Größenklassen der Debitoren und Ermittlung ihres Gesamtwerts

Deming¹⁾ beschreibt in einem Beispiel eine „Schnellprüfung von Konten in Hauptbüchern“, deren Zweck die Ermittlung der rationellsten, d. h. *ergiebigsten Einzugsmöglichkeit von Forderungen* war. Dabei ist der Debitorenbestand nicht nur nach besonderen Merkmalen (Besetzung und Gesamtbetrag bestimmter Größenklassen der Einzelforderungen) analysiert, sondern auch der Gesamtgröße nach bestimmt worden.

Folgende Einzelheiten seiner Darstellungen seien hier wiedergegeben:

Die Untersuchung betraf 23 große Bände mit je etwa 500 Seiten, von denen jede wiederum 27 Zeilen aufwies. Viele der Zeilen enthielten Namen. Sie bildeten die Debitorenliste. Bei einer Anzahl Namen waren Geldbeträge verzeichnet (z. B. \$ 752), was anzeigte, daß dieses Konto am Tage der Eintragung noch nicht ausgeglichen war.

Als *Auswahleinheit* wurden je zwei gegenüberliegende Seiten (die linke und rechte Seite eines aufgeschlagenen Bandes) definiert. Da eine Seite 27 Zeilen enthielt, bestand die Auswahleinheit aus 54 Zeilen und konnte infolgedessen bis zu 54 Namen und bis zu 54 Geldbeträge aufweisen. Die Geldbeträge stellten die *Grundgesamtheit* dar. Die Zahl der Auswahleinheiten (Doppelseiten) je Band wurden in einer „Liste der Auswahleinheiten nach Bänden“ eingetragen und mit fortlaufenden Ordnungszahlen versehen. Die Liste bildete die *Grundlage für die Stichprobenauswahl*.

Aufgrund dessen wurde z. B. in Band 1 eine Zahl von 217 Auswahleinheiten (Doppelseiten), in Band 2 eine Zahl von 224 Auswahleinheiten, in Band 3 eine Zahl von 289 Auswahleinheiten festgestellt und in jedem Band mit 1 beginnend fortlaufend numeriert. In der „Liste der Auswahleinheiten nach Bänden“ erhielten ferner die 217 Auswahleinheiten des 1. Bandes die Seriennummern 001 bis 217, die 224 Auswahleinheiten des 2. Bandes die Seriennummern 218 bis 441, die 289 Auswahleinheiten des 3. Bandes die Seriennummern 442 bis 730 usw. bis der 213. Auswahleinheit des 23. Bandes die Seriennummer 5 380 zugeteilt war.

Sodann mußten die *Wertklassen* definiert werden. Sie wurden nach den spezifischen Bedürfnissen festgelegt und hätten auch für eine Vollerhebung gegolten. Neben den Wertklassen waren zu Kontrollzwecken auch Klassen für

- a) nur einen Namen enthaltende Zeilen (ausgegliche Konten) und
- b) leere Zeilen

zu bilden.

¹⁾ (1960), S. 110 ff.

Vor der Entscheidung über die hier gewählte Definition der Auswahleinheiten war es wesentlich, sich zu vergewissern, daß in allen 23 Bänden keine Auswahleinheit einen großen Anteil von Geldbeträgen der Klassen A (Einzelbeträge von \$ 2 000 und mehr) oder B (Einzelbeträge zwischen \$ 1 000 und \$ 1 999) aufweisen konnte. Hätte diese Möglichkeit bestanden, wären die Ergebnisse bei dem hier gewählten Stichprobenplan beeinträchtigt worden. Mit Hilfe des für die Eintragungen in die Bücher Verantwortlichen wurde jedoch festgestellt, daß es völlig unmöglich sei, daß eine Auswahleinheit mehr als einige wenige Geldbeträge von den Größenordnungen der Klassen A und B aufweisen könne. Überdies ergab eine Vorprüfung der Doppelseiten in allen Bänden keinen Hinweis dafür, daß Häufungen der Klassen A oder B vorlagen.

In einer Einfügung führt *Deming* aus, daß es ein leichtes gewesen wäre, ausgezeichnete *Alternativpläne* zu entwerfen, die trotz Häufungen von Geldbeträgen der Größenklassen A und B in einzelnen Auswahleinheiten gute Resultate geliefert haben würden, aber diese Alternativpläne hätten alle mehr Geschicklichkeit und Zeit gefordert, als die Zeitbeschränkung erlaubte. Man hätte z. B. eine Zufallszahl für die Auswahl der Seite sowie zwei oder drei Zufallszahlen für die Auswahl der hierauf befindlichen Zeilen ziehen können. Eine andere Möglichkeit hätte darin bestanden, geschichtete Stichproben zu benutzen. Zum Beispiel hätte man jede Doppelseite oder eine große Stichprobe aus den ungefähr 6 000 Doppelseiten durchsehen können, um auf einer getrennten Liste die Seiten zu notieren, die mehr als z. B. 4 Eintragungen der Klasse A aufwiesen. Diese Liste hätte man dann gesondert stichprobenweise prüfen können. Wie verlockend auch jeder solche Plan gewesen sei, stellt *Deming* fest, er hätte nur zu einem Mißergebnis geführt, da weder Zeit noch Kräfte zur Verfügung gestanden hätten, ihn auszuführen.

Berechnungen hatten ergeben, daß 200 Auswahleinheiten ein *hinreichend sicheres und genaues Ergebnis* liefern würden. Die Stichprobe sollte das Vorhandensein von Konten der Klasse A (Einzelbeträge von \$ 2 000 und darüber) aufdecken und eine grobe Messung ihres Anteils unter allen noch offenen Konten bringen. Es wurde vermutet, daß der Anteil der offenen Konten der Klasse A der Anzahl nach nur etwa 1% ausmachte, dem Wert nach aber sehr bedeutend zu sein versprach.

Die *Stichprobenplanung* ging davon aus, daß die Konten der Klasse A über alle Zeilen gleichmäßig, d. h. ohne Häufungen, gestreut waren und daß die Hälfte aller Zeilen Geldbeträge aufwies. Da eine Auswahleinheit (Doppelseite) 54 Untersuchungseinheiten (Zeilen) umfaßte, betrug die in der Stichprobe zu erwartende Zahl der zur Klasse A gehörenden Zeilen aufgrund der vorstehenden Feststellungen und Annahmen $(0,5 \times 200 \times 54) \times 0,01 = 54$. Der *relative Fehler* dieser Schätzung¹⁾ betrug mit über 68% Wahrscheinlichkeit nicht mehr als 15%. Dazu wurde fest-

¹⁾ Der relative Fehler ist das Verhältnis des absoluten Wertes von e zum Schätzwert. *Deming* hat ihn mit Hilfe der *Poissonschen* Formeln (s. Formel 13, S. 26) berechnet.

Aus $e_r = \frac{e}{q} = +t \sqrt{\frac{p}{nq}}$ und $\frac{p}{nq} \approx \frac{1}{m}$ für $m \ll n$ folgt bei $t = 1$ ein $e_r \approx \sqrt{\frac{1}{m}}$ und

wegen $\sqrt{\frac{1}{m}} = \sqrt{\frac{1}{54}} = 0,15$ schließlich $e_r \approx 0,15$.

gestellt, daß die Ergebnisse auch dann noch brauchbar sein würden, wenn sich wegen einiger Ballungen der Schätzfehler als doppelt so groß erweisen sollte. Tatsächlich wurden in der Stichprobe 64 Zeilen der Klasse A festgestellt und es ergab sich, wie aus der unten wiedergegebenen Tabelle 20 zu entnehmen ist, ein relativer Schätzfehler von 30% für die Klasse A. Das bedeutete, daß die vermutliche Zahl von Zeilen der Klasse A, d. h. von Debitoren mit einem Einzelwert von mindestens \$ 2 000, mit einem Sicherheitsgrad von über 68% zwischen 1 257 und 2 187 liegen würde. Rechnet man mit einem durchschnittlichen Einzelwert dieser Debitoren von \$ 2 500, wie *Deming* das getan hat, so ergibt sich ein Gesamtbetrag, der mit 68% Sicherheit zwischen $1\,257 \times \$\,2\,500 = \$\,3\,142\,500$ und $2\,187 \text{ mal } \$\,2\,500 = \$\,5\,467\,500$ liegt.

Der Stichprobenplan sah die Prüfung der 200 Auswahlseinheiten in 10 *Unterstichproben* von je 20 Auswahlseinheiten vor.

Gleichzeitig waren die *Tabellen* zu entwerfen, die die einzelnen Untersuchungsergebnisse aufzunehmen, und weitere Tabellen, die die Ergebnisse der ersten Tabellen in geeigneter Weise zusammenzufassen hatten. Damit waren die Vorbereitungen der Erhebung abgeschlossen und die eigentlichen Aufnahmen konnten beginnen. In die so vorbereiteten Tabellen waren die Zeilenzahlen der einzelnen Stichprobeneinheiten nach Klassen gegliedert einzutragen.

Deming schildert dann Anweisungen für die *Auswertung der Tabellen und die Entwicklung der Endergebnisse* der Untersuchung aus den einzelnen Tabellergebnissen. Auf diese Ausführungen soll hier nicht näher eingegangen werden, da die darin enthaltenen Anweisungen und Tabellen über den Rahmen dieser Untersuchung hinausgehen.

Zur Vervollständigung der Darstellung dieses Beispiels sei das *Ergebnis der Untersuchung* in Form folgender Tabelle wiedergegeben.

Tabelle 17: Geschätzter Umfang und Gesamtwert je Größenklasse und insgesamt unter Zugrundelegung eines Sicherheitsgrades von 68,3 % ($t = 1$)

Größenklasse in \$	Betrag in \$ 1000	Geschätzte Zahl der Konten		
		absolut	in % aller Konten*)	in % der Konten mit Saldo*)
Konten insgesamt	20 309	223 614		
Konten ohne Salden		68 750	$30,3 \pm 1,3$	
Konten mit Salden	20 309	154 864	$69,7 \pm 1,3$	
2 000 und mehr	4 310	1 722		$1,1 \pm 0,3$
1 000 bis 1 999,99	2 692	1 802		$1,2 \pm 0,2$
400 bis 999,99	3 785	5 407		$3,5 \pm 0,4$
100 bis 399,99	5 064	20 256		$13,1 \pm 0,8$
50 bis 99,99	1 758	23 457		$15,1 \pm 0,6$
20 bis 49,99	2 350	67 223		$43,4 \pm 1,9$
unter 20	350	34 997		$22,6 \pm 1,5$

*) Die Zahlen nach dem Doppelzeichen ($\pm$) beziffern die geschätzten Standardfehler.

Da die Arbeit an einem Wochende, d. h. zwischen Dienstschluß am Freitagabend und Dienstbeginn am folgenden Montagmorgen, durchgeführt sein mußte, wurde die Auswahlseinheit nicht nach Maßgabe der größten Wirksamkeit des Stichprobenplans, sondern mit Rücksicht auf die *zeitliche Begrenzung der Untersuchung* bestimmt.

Für die Arbeiten standen drei Helfer zur Verfügung; der erste, ein erfahrener Fachmann, der Zweite ein Neuling in dieser Tätigkeit und eine dritte Kraft, von der der Mandant sagte, sie könne die Auswahlseinheiten im vorhinein zählen und die dicken Bände je nach Bedarf von einem Tisch auf den anderen schaffen.

Das Beispiel zeigt, daß bei Anwendung selbst relativ einfacher Verfahren bereits *beachtlich detaillierte und genaue Ergebnisse in kurzer Zeit* zu erzielen sind. Die umfangreichen Voruntersuchungen können, wie *Deming* bemerkt, bei Wahl eines anderen, verfeinerten Stichprobenplans, zugunsten einer besseren Hauptuntersuchung abgekürzt werden, die dann zugleich den Vorteil bietet, noch genauere Ergebnisse, d. h. noch kleinere Vertrauensbereiche zu liefern.

Beispiel 2

Kreditfristanalyse in Verbindung mit einer Gesamtwertschätzung der Debitoren

Ein anderes Beispiel einer Bestandsanalyse beschreiben *Trueblood/Cyert*¹⁾ in einem „*Typischen Bericht über eine Kreditfristanalyse bei den XYZ-Warenhäusern*“. Es handelt sich um die Schätzung der erforderlichen Wertberichtigung für einen Debitorenbestand mit gleichzeitiger Analyse der Debitoren nach Kreditfristen und Schätzung des auf die einzelnen Fristenklassen entfallenden Anteils der Gesamtforderung. Die Ergebnisse sind in der folgenden Tabelle zusammengefaßt.

Tabelle 18: Schätzung der erforderlichen Wertberichtigung und der Zusammensetzung des Debitorenbestandes nach Kreditfristen

Kreditfrist	Stichproben- anteil dieser Klasse in %	Geschätzter Wertanteil dieser Klasse am Gesamtwert in \$ 1 000	Vertrauensbereich bei einem Sicherheitsgrad von 95%			
			in % des Schicht- wertes	in \$ 1 000	Bereichs- grenze	
					untere	obere
					in \$ 1 000	
täglich fällig	78,05	4 040	± 1,4	± 57	3 984	4 097
3 Monate	7,99	414	± 6,4	± 26	387	440
4 Monate	7,84	406	± 8,9	± 36	370	442
5 Monate	4,54	235	± 10,1	± 24	211	259
6 Monate	0,96	50	± 21,6	± 11	39	60
7—12 Mon.	0,62	32	± 24,7	± 8	24	40
Auswahlsatz 15%						
Geschätzter Gesamtwert		5 176				
Geschätzte erforderliche Wertberichtigung		150	± 3,67	± 5,5	145	156

Mit einer Stichprobe, die 15% der Debitorenkonten umfaßte, wurde als wahrscheinlich erforderliche Wertberichtigung ein Betrag von ca. \$ 150 000 geschätzt. Der Wertberichtigungsbetrag, der bei lückenloser Prüfung aller Debitorenkonten ermittelt worden wäre, hätte sich mit einem Sicherheitsgrad von 95% um nicht mehr als ca. \$ 5500, d. s. 3,67% dieses Schätzwertes, von ihm unterschieden und damit zwischen \$ 144 500 und \$ 155 500 gelegen.

Mit der *gleichen Stichprobe* wurde die Fristigkeit der Debitorenkonten untersucht. Dabei ergab sich, daß der Vertrauensbereich für den Gesamtwert der einzelnen Schichten, gemessen an diesem Schichtenwert, von ± 1,4% in der Klasse mit dem

1) a. a. O., S. 122 ff.

größten Anteil von 78% der Einheiten bzw. des gesamten Debitorenwertes bis auf 24,7% in der Klasse mit dem kleinsten Anteil (rd. 0,6%) der Einheiten bzw. des Wertes anstieg.

Daß die Genauigkeit der Schätzung mit geringerem Umfang der einzelnen Kreditfristklassen immer geringer, d. h. der Vertrauensbereich bei gleichem Sicherheitsgrad für alle Klassen immer größer wurde, beruht darauf, daß die Planung der Stichprobe in erster Linie auf die Genauigkeit der *Schätzung der Wertberichtigung* abgestellt war.

2. Prüfung der Geschäftsvorfälle

a) Prüferische Zielsetzungen und statistischer Charakter der Prüfungsfragen

Obwohl das Schwergewicht der Prüfung heute insbesondere bei Großbetrieben auf der Prüfung des innerbetrieblichen Kontrollsystems und der Bestände liegt, muß noch untersucht werden, ob die Prüfung der *materiellen Richtigkeit der Verbuchung der Geschäftsvorfälle* mit Hilfe statistischer Methoden erfolgen kann.

Nach den Grundsätzen ordnungsmäßiger Buchführung sind die buchhalterisch wesentlichen Daten eines Geschäftsvorfalles in einem Beleg zusammenzufassen, der als Grundlage für die Verbuchung der Daten in den Konten dient. Der Prüfer muß sich daher an Hand der Originalunterlagen vergewissern über

1. die Übereinstimmung des Beleges mit dem Geschäftsvorfall,
2. die Übereinstimmung der Buchung mit dem Beleg.

Bei dieser Prüfung ist festzustellen, ob die Geschäftsvorfälle

1. rechnerisch richtig und
2. entsprechend ihrem Inhalt auf den zutreffenden Konten erfaßt sind.

In aller *Regel* verändern Verstöße gegen diese Forderungen den Inhalt der Bilanz oder der Gewinn- und Verlustrechnung dem Betrag oder der Gliederung nach. Es handelt sich daher hier um *materielle Erfordernisse* im Sinn der in der vorliegenden Arbeit verwandten Terminologie.

Nur in einigen *Sonderfällen* (nämlich bei Unterkonten) brauchen derartige Verstöße nicht unmittelbar eine Auswirkung auf den Jahresabschluß zu haben. Es ist möglich, daß ein Zahlungseingang oder -ausgang auf einem falschen Personenkonto verbucht wird, z. B. nicht auf dem Konto Meier, sondern auf dem Konto Meyer, dann stimmen die Personenkonten nicht, trotzdem weist das Sachkonto Debitoren und folglich auch die Bilanz den richtigen Saldo aus. Oder es wird eine Frachtenzahlung nicht auf dem Unterkonto Bahnfrachten, sondern irrtümlich auf dem Unterkonto Autofrachten verbucht; auch dann ist der Ausweis der Frachten insgesamt richtig.

Obwohl die genannten Fehler dann den Inhalt des Jahresabschlusses nicht unmittelbar zu beeinflussen brauchen, tun sie es häufig mittelbar. Kommen z. B. Verwechslungen auf den Personenkonten vor, so weisen einzelne Personenkonten höhere Salden aus, als der Höhe der wirklichen Forderung entspricht, andere zu niedrige. Das kann zur Folge haben, daß die Zahlung der unberechtigten Mehrforderung abgelehnt, die nicht in Rechnung gestellten Forderungen aber gleich-

falls nicht bezahlt werden. Auch bei Verwechslung von Unterkonten eines Sachkontos kann eine Auswirkung auf die Bilanz eintreten. Sind z. B. die Bahnfrachten Teile der Herstellungskosten von bestimmten Fertig- oder Halbfertigwaren, so verfälschen die nicht erfaßten oder zuviel erfaßten Bahnfrachten den Bestandswert dieser Fertig- oder Halbfertigwaren, während die Frachten in der Gewinn- und Verlustrechnung insgesamt richtig ausgewiesen sind. Insofern haben diese Fehler also ebenfalls materielle Bedeutung.

Die Prüfung dient aber nicht nur der Feststellung der Richtigkeit der Verbuchung der einzelnen Geschäftsvorfälle. Es soll vielmehr aus den beobachteten Einzelfällen ein *Urteil über das Gesamtergebnis* gewonnen, nämlich festgestellt werden, ob die in den Jahresabschluß eingegangenen Salden den gesetzlichen Vorschriften hinsichtlich Bewertung und Ausweis entsprechen. Im Falle der Feststellung von Fehlern soll der Gesamtfehler ermittelt werden, um die Abschlußposition zu berichtigen.

Unter Umständen müssen auch *Mittelwertbestimmungen* und *Aufgliederungen der Konteninhalte* vorgenommen werden, um angemessene Wertansätze und Prüfungsurteile zu finden.

Wie schon mehrfach ausgeführt, genügt zum Erreichen dieser Prüfungsziele die Ermittlung von Näherungswerten. Dementsprechend lassen sich diese Fragen grundsätzlich mit Hilfe von *Stichprobenprüfungen* beantworten, wobei Schätzungen sowohl von Gesamt- oder Mittelwerten wie von Anteilen in Betracht kommen.

Schätzungen von *Gesamtwerten* oder *Mittelwerten* kommen z. B. für den Schluß von den geprüften Einzelbelegen und -buchungen auf die Richtigkeit des Wertansatzes der Bilanzposition oder für die Ermittlung von Durchschnittswerten, wie z. B. des durchschnittlichen Zinssatzes verschiedener Mietobjekte oder des durchschnittlichen Abschreibungssatzes für verschiedene Produktionsmittel, in Frage.

Schätzungen von *Anteilen* sind zweckmäßig, um z. B. bei Kontenverwechslungen deren Wertanteil am Gesamtwert des Kontos zu ermitteln und um hiernach den betreffenden Bilanzausweis zu berichtigen. Sind z. B. irrtümlich Gehälter über Lohnkonto verbucht, so genügt es im allgemeinen den prüferischen Bedingungen, wenn die Falschbuchungen global berichtigt werden. Anteilsschätzungen können ferner zur Analyse von Konten benutzt werden, z. B. wenn es darauf ankommt, ein Sammelkonto „Frachten“ in die Bestandteile Auto-, Bahn-, See-, Luftfrachten usw. aufzuteilen.

b) Schätzung von Gesamtwerten, Mittelwerten und Anteilen mittels Schätzstichproben

Der diesen Stichprobenverfahren zugrunde liegende *Formelapparat* ist, soweit es sich um die Schätzung von *Gesamtwerten* handelt, auf S. 58 ff., soweit es sich um die Schätzung von *Mittelwerten* handelt, auf S. 74, und soweit es sich um die Schätzung von *Anteilen* handelt, auf S. 40 ff. erörtert worden.

Infolgedessen kann auch hinsichtlich der *Wirkungen der Faktoren* in diesen Formeln auf den jeweils erforderlichen Stichprobenumfang auf die entsprechenden Ausführungen (s. S. 59 f., 74, 44 ff.) verwiesen werden.

c) *Anwendbarkeit auf das Prüfungsgebiet, insbesondere Prüfbarkeit der unter Einsatz selbsttätig arbeitender Geräte erstellten Rechnungslegung mittels Stichproben*

(1) *Prüferische Auswertung*

Im Hinblick auf die Abhängigkeit der Genauigkeit des Stichprobenergebnisses von der Homogenität der Grundgesamtheit bietet die stichprobenweise Prüfung solcher Konten den größten Rationalisierungserfolg, deren Inhalt sich durch *großen Umfang* und starke wertmäßige Gleichförmigkeit, d. h. *kleine Variabilität*, auszeichnet. Diese Bedingungen sind mit zunehmender Größe der Betriebe in immer stärkerem Maße gegeben. Z. B. liegt bei der Lohn- oder Materialbuchführung, aber auch bei der Buchführung über Warenausgangsrechnungen, Frachten, Zinsen usw. oft ein sehr umfangreicher Buchungsstoff vor, dessen Einheiten inhaltlich jeweils gleichartig sind und die dem Betrag nach in bekannten oder schnell bestimmbar Grenzen liegen.

In diesen Fällen ist die Auswahl einer Stichprobe aus den geordnet abgelegten Belegen leicht möglich. An Hand der Belege kann einerseits der zu Grunde liegende Geschäftsvorfall nach prüferischen Maßstäben untersucht, andererseits aus dem Ergebnis der Stichprobe auf den Gesamtumsatz des Kontos bzw. dessen Saldo geschlossen werden. Mit Hilfe dieses Schätzwertes kann dann entweder die Richtigkeit der Abschlußposition bestätigt oder die Größe des vermutlichen Fehlers mit der gewünschten Genauigkeit bestimmt werden.

Damit erweist sich die *Zufallsstichprobe als ein für die Prüfung großer Buchungsmassen besonders geeignetes Prüfungsverfahren*, weil sie einen Schluß von einer in der Regel relativ kleinen Anzahl von Einzelbelegen, der Stichprobe, unmittelbar auf das Gesamtergebnis des Kontos erlaubt, ohne daß eine Prüfung der Belegübernahme in die Bücher, der Additionen, der Überträge usw. erforderlich wird.

Diese Feststellung kann besondere Bedeutung für die Prüfung von *Buchführungen, bei denen selbsttätig arbeitende Geräte eingesetzt sind*, erlangen. Die Einführung der Automation in die Buchführung begegnet vom prüferischen Standpunkt aus einigen Vorbehalten. So schreibt Finanzpräsident *Schröder*, Bonn, über „Die Auswirkungen des Einsatzes von datenverarbeitenden Maschinen in den Unternehmungen auf die steuerliche Betriebsprüfung“¹⁾:

„Die mit der neuartigen elektronischen Buchungstechnik erstellten Rechnungswerke sind jedoch in manchen Fällen nicht frei von Bedenken wegen der Ordnungsmäßigkeit der Buchführung. Die Gefahr, Grundsätze ordnungsmäßiger Buchführung außer acht zu lassen, ist deshalb größer als bei der mechanisierten Buchführung, weil man aus Gründen der Kostenersparnis den Weg von der Eingabe bis zum Ergebnis nicht mehr voll überschaubar macht. Es fehlt an der notwendigen Übersichtlichkeit, ja Überprüfungen sind für gewisse Gebiete des Rechnungswesens erschwert und in Einzelfällen ... sogar unmöglich.“

Schröder verlangt daher zum Zweck der Nachprüfbarkeit eine Aufzeichnungspflicht nicht nur der Geschäftsvorfälle in den Belegen, sondern auch der Buchungsanweisungen und Programmierungen. Er schreibt²⁾:

1) In „Die steuerliche Betriebsprüfung“, Jg. 1, 1961, Heft 1, S. 4.

2) a. a. O., S. 7.

„An die Mitwirkungspflicht des Steuerpflichtigen zur Aufklärung sind höhere Anforderungen zu stellen als früher bei der manuell geschriebenen Buchführung. Über die Vorlage aller die Buchführung betreffenden Organisations- und Arbeitspläne, Buchungsanweisungen und Programmierungen hinaus müssen schriftlich festgehaltene Vorgänge in Protokollen, Niederschriften und sonstigen Unterlagen zur Verfügung gestellt werden. Nur im Klartext geschriebene Unterlagen sind für die Überprüfung ausreichend, und Zahlenverdichtungen müssen durch Tabellen und Auflistungen so aufgegliedert werden, daß von der Buchung bis zum Beleg und vom Beleg bis zur Buchung der Geschäftsvorgang verfolgt und die Objektbezogenheit durch entsprechende Belege nachgewiesen werden kann.“

Mit dem Problem der Sicherung der Ordnungsmäßigkeit kaufmännischer Rechnungslegung beim Einsatz selbsttätig arbeitender Geräte hat sich auch das deutsche Institut der Wirtschaftsprüfer befaßt und ähnliche Forderungen wie *Schröder* erhoben. Diese Anforderungen hat es kürzlich in seiner Fachzeitschrift „Die Wirtschaftsprüfung“¹⁾ in einer Stellungnahme zu den „Mindestanforderungen zur Sicherung der Ordnungsmäßigkeit kaufmännischer Rechnungslegung beim Einsatz selbsttätig arbeitender Geräte“ veröffentlicht. Hierin werden Erfordernisse bezüglich des Nachweises der eingegebenen Daten, der Sicherung des programmgemäßen Ablaufs, der Kontrolle der Rechnungsergebnisse und des innerbetrieblichen Kontrollsystems aufgestellt. Diese Erfordernisse sollen nicht zuletzt die Nachprüfbarkeit der Vorgänge sicherstellen und erleichtern.

Auch bei Vorlage der verlangten Aufzeichnungen über Problembeschreibungen, der Ablaufdiagramme sowie der verwandten Programme ist keineswegs sicher, daß der Prüfer tatsächlich von allen Programmänderungen, insbesondere solchen, die durch Bedienungseingriffe herbeigeführt worden sind, erfährt. Irrtümliche oder absichtliche Falscheingaben, mangelnde Kontrollschaltungen, Maschinendefekte bleiben als Fehlerquellen bestehen, so daß zur Nachprüfung unter Umständen Arbeitswiederholungen erheblichen Ausmaßes erforderlich werden. In jedem Fall werden wesentlich höhere Anforderungen an die technischen Kenntnisse des Prüfers gestellt, als das bei der Prüfung der konventionellen und Lochkarten-Buchführungsverfahren der Fall ist.

Das *Zufallsstichprobenverfahren* bietet aber einen von den Geräten, ihren Programmen sowie von den Kontrolleinrichtungen und -aufzeichnungen völlig unabhängigen zweiten Prüfungsweg, denn es erlaubt, von den in die Stichprobe gezogenen Belegen und den ihnen zugrunde liegenden Geschäftsvorfällen unter Überspringen der Prüfung der Belegübernahme in die Maschine, der Programmierung und der Arbeitsweise der Maschine unmittelbar auf das Ergebnis der Rechnungslegung zu schließen und einen Vertrauensbereich anzugeben in dem dieses Ergebnis mit einem vom Prüfer nach fachlichem Ermessen bestimmbaren Sicherheitsgrad liegt.

Zu dem gleichen Ergebnis kommen *Davis/Rounsaville*, wenn sie schreiben²⁾:

„Mit der derzeitigen schnellen Einführung der automatischen datenverarbeitenden Maschinen durch die öffentliche Verwaltung und die Großbetriebe in ganz Amerika stehen jetzt weite Kreise des Prüferberufs (externe und interne Prüfer)

¹⁾ a. a. O., Jg. 14, Nr. 11, S. 308.

²⁾ *Davis, D. D./Rounsaville, A.*, Auditing Voluminous Data by Modern Sampling Methods, JofA, Juni 1959, S. 46.

dem Problem der Prüfung buchstäblich von Bergen durch die Automation gelieferter Daten gegenüber. Statistische Stichproben sind ein Teil der Lösung. Diese Daten sind nicht nur sehr zahlreich, sondern auch in großem Umfang ähnlich oder gleich. In vielen Fällen stimmen die Urdokumente in diesen beiden Punkten überein. Dadurch eignen sie sich für die Anwendung von Zufallsstichproben. In vielen Fällen können die datenverarbeitenden Automaten selbst dazu benutzt werden, die Auswahl der Zufallsstichproben zu unterstützen.“

(2) Grundlagen für die Festsetzung der Bestimmungsfaktoren Umfang der Grundgesamtheit und Varianz in ihr

Ähnlich wie bei den Gesamtwertschätzungen sind die *Grundgesamtheiten* jeweils diejenigen Geschäftsvorfälle bzw. ihre Dokumentationen, deren Ergebnisse entsprechend den gesetzlichen Vorschriften in einer Position des Jahresabschlusses gesondert auszuweisen sind. Die *Einheiten* sind im wesentlichen die Buchungen auf den Sachkonten, den Journalen oder die Belege. Welche dieser Möglichkeiten im Einzelfall gewählt wird, richtet sich nach der vorgesehenen Prüfungsrichtung (progressiv — retrograd). *Untersuchungsmerkmal* ist in der Regel der Geldbetrag des einzelnen Geschäftsvorfalles, es kann aber auch die Menge (kg, m³ etc.) sein. Der *Umfang* der Grundgesamtheit und die *Varianz* der Merkmalwerte sind gleichfalls nach den oben (S. 50 ff., S. 61 ff.) besprochenen Grundsätzen zu ermitteln.

Auswahlgrundlagen und Auswahlverfahren

Es ist nun zu untersuchen, ob und wo unter den Dokumentationen der Geschäftsvorfälle eine *Ordnung* besteht, die als *Auswahlgrundlage* geeignet ist.

Die *Schriften* über die Geschäftsvorfälle sind häufig nach anderen als rein buchführerischen Gesichtspunkten geordnet, z. B. alphabetisch, zentral oder dezentral, nach Abteilungen, Sachbearbeitern, die grundlegenden Geschäftsvorfälle in Grundakten, die hierzu gehörenden laufenden Vorgänge in einer Akte für die laufenden Geschäftsvorfälle usw. Häufig kommen die verschiedensten Ablageordnungen nebeneinander vor. In diesen Fällen ist es schwierig und zeitraubend und daher kostspielig, eine die vorerwähnten Bedingungen erfüllende Ordnung der Geschäftsvorfälle herzustellen.

In der Regel ist eine solche Ordnung in der Ablage der *Belege* vorhanden. Sie sind

- a) nach der zeitlichen Reihenfolge und/oder
- b) nach sachlichen Gesichtspunkten

geordnet aufbewahrt.

Doch braucht auch dies nicht immer zuzutreffen, weil die Belege auch in den jeweiligen Akten der Debitoren oder Kreditoren abgelegt werden. Solche Ordnungsprinzipien erschweren ein auf den Belegen als Ausgangspunkte beruhendes Prüfungsverfahren.

Dagegen sind die Buchungen stets nach

- a) der zeitlichen Reihenfolge in den Grundbüchern bzw. Journalen und
- b) nach sachlichen Gesichtspunkten in den Sachkonten

geordnet.

Häufig fallen *beide Ordnungsprinzipien* ganz oder teilweise zusammen, z. B. im amerikanischen Journal, das die Geschäftsvorfälle gleichzeitig chronologisch und sachlich ordnet, oder die einzelnen Geschäftsvorfälle werden lediglich in Sach-

konten (hier in zeitlicher Reihenfolge) verbucht; dies ist insbesondere bei Belegbuchhaltungen der Fall, bei denen ohne Benutzung von Journalen vom Beleg direkt auf das Sach- (bzw. Personen-)konto gebucht wird. Die Buchungen im Sachkonto erfolgen dabei häufig in listenmäßiger Zusammenfassung des Buchungsstoffes.

Es ist unmöglich, hier auf alle verschiedenen Buchführungsverfahren, wie z. B. Bücherbuchführungen, Loseblattbuchführungen, manuelle und maschinelle Durchschreibeverfahren, Lochkartenbuchhaltung oder Verfahren mit elektronischer Datenverarbeitung und auf die verschiedenen Belegablageverfahren einzugehen, zumal alle diese Verfahren je nach der Unternehmensorganisation im Einzelfall sehr häufig Besonderheiten aufweisen.

Eine der vorgenannten Ordnungen muß gegeben sein, wenn die Buchführung des Unternehmens als ordnungsmäßig bezeichnet werden soll. Es liegt daher in jedem Fall entweder in den Buchungen oder den Belegen, häufig sogar in beiden eine *Ordnung* vor, die eine *Zufallsauswahl* ermöglicht.

Je nach der im Einzelfall vorgefundenen Ordnung kann die Prüfung — ebenso wie bei den Bestandsprüfungen — *progressiv* oder *retrograd* durchgeführt werden. Von den Konten kommen die *Journale* als Auswahlgrundlage praktisch nicht in Frage, weil im Journal die Geschäftsvorfälle in chronologischer Reihenfolge aufgezeichnet werden. Das hat sehr ungleichmäßige Verteilungen der Buchungen der zu den verschiedenen Abschlußpositionen gehörenden Geschäftsvorfälle zur Folge. So treten z. B. Häufungen von Lohn-, Gehalts- und sonstigen regelmäßigen Zahlungen in bestimmten periodischen Abständen, Häufungen von Abschreibungs- bzw. Wertberichtigungs- sowie Rückstellungsbuchungen usw. jeweils kurz vor Jahresabschlüssen auf. Hinzu kommt, daß Abschreibungs- bzw. Wertberichtigungs-, Rückstellungs- und ähnliche Buchungen einen so geringen Anteil aller Geschäftsvorfälle darstellen, daß eine Analyse nach allen Abschlußpositionen nur mit einer unverhältnismäßig großen Stichprobe hinreichend genau und zuverlässig möglich ist.

In den *Sachkonten* ist der Prüfungsstoff demgegenüber grundsätzlich nach Gliederungsgesichtspunkten geordnet, die für den Ausweis der Geschäftsvorfälle im Jahresabschluß von Bedeutung sind. Häufig kommt es vor, daß Sachkonten, die einer Position des Jahresabschlusses entsprechen, noch unterteilt sind, z. B. das Konto „Zinsaufwand“ in die Unterkonten Bank-, Wechsel-, Anleihezinsen usw. In solchen Fällen sind verschiedene *Auswahlverfahren* möglich.

Soll ein *Sammelkonto*, das sich aus mehreren Unterkonten zusammensetzt, geprüft werden, so bieten sich drei verschiedene Möglichkeiten für die Auswahl der Stichprobeneinheiten:

- a) Der Prüfer betrachtet die auf den einzelnen Unterkonten vorgenommenen Buchungen als jeweils getrennte Grundgesamtheiten und prüft diese getrennt.
- b) Der Prüfer kann aber auch alle Buchungen aller Unterkonten als eine einzige Grundgesamtheit ansehen und bestimmt die Stichprobeneinheiten dann entweder
 - i) durch ein einfaches Auswahlverfahren von allen Unterkonten gemeinsam oder
 - ii) durch eine nach Unterkonten geschichtete Auswahl.

Werden die *Unterkonten* als getrennte Grundgesamtheiten betrachtet, so kann durch Gesamtwertschätzungen auf das Endergebnis (Umsatz, Saldo) jedes Unterkontos geschlossen werden.

Bilden *alle Buchungen* aller Unterkonten die Grundgesamtheit, so kann bei einfacher Zufallsauswahl durch eine Gesamtwertschätzung eine Aussage über die Ergebnisse des Sammelkontos gemacht werden, nicht aber ohne weiteres über die Ergebnisse der einzelnen Unterkonten. Wird auch über sie Auskunft verlangt, so bedarf es einer Analyse der Stichprobe um festzustellen, welche Unterkonten von den in die Stichprobe gekommenen Buchungen bzw. Geschäftsvorfällen berührt sind und wie sich diese Buchungen und der Gesamtwert ihrer Buchungsbeträge auf die Unterkonten verteilen. Bei der Hochrechnung dieser Ergebnisse auf die Grundgesamtheit mittels Anteilsschätzungen können die bereits bei der Untersuchung der analytischen Prüfungen von Beständen dargestellten Probleme auftreten (z. B. unterschiedliche Genauigkeiten der Schätzwerte für die einzelnen Anteile, s. S. 74 f.).

Berücksichtigt man aber die gegebenen *Kontenuntergliederungen* bereits bei der Auswahl der Stichprobeneinheiten, indem man die Unterkonten als Schichten betrachtet, so kann das Aussageergebnis der Stichprobe detaillierter, die Stichprobe selbst also wirkungsvoller werden. Denn die Berücksichtigung dieser Gliederung bei der Stichprobenauswahl erlaubt nicht allein die Prüfung des Gesamtergebnisses, also des Sammelkontos, sondern auch die rationelle Ermittlung der interessierenden Merkmalwerte der Schichten, also der Umsätze bzw. Salden der Unterkonten. Bei genügender Größe des Stichprobenumfangs einer Schicht (Unterkonto) ist außerdem im Bedarfsfall eine zusätzliche Analyse des Unterkontos nach weiteren Einzelheiten (z. B. bei Autofrachten, Nahverkehr, Fernverkehr) an Hand der Stichprobe möglich.

Es kann aber auch der Fall eintreten, daß nur ein bestimmter *Teil* des in einem Konto erfaßten Buchungsstoffs näher interessiert (für den kein eigenes Unterkonto geführt wird). In diesem Fall ist es schwierig, eine Auswahlgrundlage zu erstellen, die es ermöglicht, nur interessierende Buchungen in die Stichprobe zu ziehen (z. B. aus dem Konto Frachten nur die Frachten für eine bestimmte Ware). Es ergibt sich dann die Frage, ob es notwendig ist, die Einheiten mit dem interessierenden Merkmal zunächst auszusondern (z. B. an Hand der Belege) oder ob die gegebene Auswahlgrundlage, d. i. das Konto, beibehalten werden kann, was dann zu einer entsprechenden Ausdehnung des Stichprobenumfangs führen muß.

Mit diesem Problem hat sich *Deming*¹⁾ befaßt. Er berechnet zunächst die Zahl von interessierenden Untersuchungseinheiten, die erforderlich ist, um das benötigte Ergebnis mit der erforderlichen Genauigkeit und dem gewünschten Sicherheitsgrad zu erhalten. Diese Berechnungen können mit Hilfe der üblichen Formeln vorgenommen werden, deren einfachste oben (s. S. 44 ff.) beschrieben worden sind. *Deming* schätzt dann den Anteil der Einheiten mit dem interessierenden Merkmal in der zu untersuchenden Grundgesamtheit. Durch Multiplikation des reziproken Wertes dieses Anteils mit der erforderlichen Zahl interessierender Einheiten bestimmt er den Umfang der zu ziehenden Stichprobe. In einem von ihm beschriebenen Beispiel der Berechnung der auf eine Eisenbahngesellschaft

1) (1960), S. 126 ff.

entfallenden Frachtenanteile von Gemeinschaftstransporten¹⁾ aus einer sehr großen Grundgesamtheit von Frachtabrechnungen hat er sich für diese Methode der Ausdehnung des Stichprobenumfangs entschieden, weil sie kostengünstiger war, als die vorherige Aussonderung der interessierenden Frachtabrechnungen.

Sicherheitsgrad und Vertrauensbereich

Im Rahmen der Erörterung der Grundlagen für die Festsetzung der Bestimmungsfaktoren muß abschließend auch noch auf den Sicherheitsgrad und den Vertrauensbereich eingegangen werden.

Bezüglich des *Sicherheitsgrades* kann auf die obigen Ausführungen (s. S. 53 ff.), insbesondere über die Bedeutung des innerbetrieblichen Kontrollsystems, verwiesen werden.

Bei den analytischen Prüfungen sollte der *Vertrauensbereich* von dem Zweck der Untersuchung und der hierfür erforderlichen Genauigkeit abhängig gemacht werden. Werden z. B. die Bahnfrachten von bestimmten Wareneingängen benötigt, um Bewertungen von Halb- und Fertigfabrikaten zu prüfen, in die diese Bahnfrachten eingehen, so hängt die Festsetzung des Vertrauensbereichs für die Analyse des Frachtenkontos von der Genauigkeit ab, mit der der Wert der Bahnfrachten bei der Bewertung der Halb- und Fertigfabrikate benötigt wird. Ist hier ein sehr genauer Wert erforderlich, so muß auch der Vertrauensbereich für den Bahnfrachtenanteil in der Analyse sehr eng bemessen werden.

Für die Gesamtwertermittlung ist der Vertrauensbereich nach den allgemeinen Grundsätzen festzusetzen, die auch für die Bestandsprüfungen gelten und auf die nunmehr eingegangen wird.

3. Festsetzung des Vertrauensbereichs und Sicherheitsgrades für Stichproben bei der Prüfung materieller Tatbestände

Bisher ist in der Literatur immer darauf hingewiesen worden, daß die Festsetzung des *Vertrauensbereichs* bei der Prüfung materieller Tatbestände grundsätzlich dem prüferischen Ermessen vorbehalten bleiben müsse. Aus dem Zweck der Abschlußprüfung lassen sich aber teilweise objektive Maßstäbe ableiten. Im Interesse des Aktionär- und Gläubigerschutzes soll das Unternehmen im Jahresabschluß darüber Auskunft geben, wie das von Aktionären und Gläubigern zur Verfügung gestellte Kapital angelegt und ob und in welchem Umfang es erhalten, vermehrt oder vermindert worden ist.

Je größer der ausgewiesene *Gewinn* sowohl absolut als auch gemessen am Eigenkapital (Grundkapital + Reserven + evtl. Gewinnvortrag) ist, um so weniger beeinflußt eine Abweichung des ausgewiesenen Gewinns vom tatsächlichen Gewinn den Zweck der Prüfung, nämlich die Sicherung des Gläubiger- und Aktionärschutzes, m. a. W. je größer der Gewinn ist, um so größer kann der Vertrauensbereich sein und umgekehrt.

Dieser Satz kann verfeinert werden, wenn man den Gesamtgewinn in den zur Verteilung an Aktionäre und Organe vorgesehenen Teil und den für die Rücklagenbildung vorgesehenen Teil *untergliedert* und den *Vertrauensbereich nur nach*

¹⁾ (1960), S. 130 ff.

der *Größe des ersten Teils* ausrichtet. Der Teil des Gewinns, der durch Zuführung zur Rücklage dem Unternehmen erhalten bleibt, ist unter dem Gesichtspunkt des Gläubiger- und Aktionärsschutzes seitens des Prüfers weniger kritisch zu betrachten als der Teil des Gewinns, der ausgeschüttet wird und damit das Unternehmen verläßt und so die materiellen Interessen der Gläubiger unmittelbar und die der Aktionäre auf die Dauer auch mittelbar beeinträchtigen kann.

Dieser Grundsatz zur Bestimmung des Vertrauensbereichs versagt jedoch, wenn der Gewinn, gemessen am Eigenkapital, sehr klein ist oder ein Verlustergebnis vorliegt.

Ist der Gewinn klein, kommt als ergänzender Maßstab die *Größe des Eigenkapitals* für die Bestimmung der Größe des Vertrauensbereichs in Betracht: Je größer das Grundkapital und die Reserven sind, desto größer kann der Vertrauensbereich ohne nennenswerte oder wesentliche Gefährdung des Grundsatzes des Aktionärs- und Gläubigerschutzes sein und umgekehrt.

Die gleiche Überlegung gilt prinzipiell auch bei der Prüfung von Jahresabschlüssen mit Verlustergebnis. Neben dem Eigenkapital oder Grundkapital kann auch die Höhe der gesetzlichen und freien Reserve als Bezugsgröße für die Festsetzung des Vertrauensbereichs herangezogen werden.

Aufgrund der vorstehenden Überlegungen erscheint der Vertrauensbereich für den Gesamtumfang der Abweichungen der Bilanzwerte vom gesetzlich zulässigen Wert und korrespondierend für die Salden der Konten der Gewinn- und Verlustrechnung *objektiver* bestimmbar, als wenn seine Festsetzung ausschließlich dem Ermessen des Prüfers überlassen bliebe.

Bei der Untersuchung, wie dieser Vertrauensbereich — er sei als „*Vertrauensbereich der Abschlußprüfung*“ (kurz: Gesamtvertrauensbereich) bezeichnet — auf die einzelnen Abschlußpositionen verteilt werden kann, kann man davon ausgehen, daß die in den einzelnen Bilanzpositionen zusammengefaßten Gesamtheiten von Bestandsposten Teile der (übergeordneten) Grundgesamtheit aller am Schluß des Geschäftsjahres vorhandenen (Aktiv- und Passiv-)Bestände des Unternehmens darstellen. Ebenso lassen sich auch die in den Positionen der Gewinn- und Verlustrechnung zusammengefaßten Gesamtheiten von Geschäftsvorfällen als die Teile der (übergeordneten) Grundgesamtheit der im Geschäftsjahr angefallenen erfolgswirksamen Geschäftsvorfälle betrachten.

Nach den Grundsätzen ordnungsmäßiger Abschlußprüfung ist jede Abschlußposition, d. h. Teilgesamtheit dieser Grundgesamtheiten, zu prüfen. Danach ist die geschichtete Stichprobe das geeignete Erhebungsverfahren. Beim geschichteten Stichprobenverfahren ergibt sich das Verhältnis zwischen dem Vertrauensbereich für die Grundgesamtheit (Gesamtvertrauensbereich) und den besonderen Vertrauensbereichen für die einzelnen Schichten nach folgender Formel:

$$(41) \quad \frac{e^2}{t^2} = \sum_{i=1}^{N'} \frac{e_i^2}{t_i^2}$$

wobei e der Meßwert für den Gesamtvertrauensbereich, t der „*Sicherheitsgrad der Abschlußprüfung*“ (kurz: Gesamtsicherheitsgrad), e_i der Meßbetrag für den besonderen Vertrauensbereich der i . Schicht (Abschlußposition) — $i = 1, 2, \dots$,

N' – und t_1 der nach der festgestellten Qualität des innerbetrieblichen Kontrollwesens in der i . Abschlußposition erforderliche Sicherheitsgrad sind.

Zu Formel (41) gelangt man aufgrund folgender Überlegung. Formel (36), Seite 59, gibt die Varianz unter den Schätzwerten für den Gesamtwert einer Grundgesamtheit im Falle eines zweistufigen Auswahlverfahrens an. In ihr ist die Schichtung als „Grenzfall“ eines zweistufigen Auswahlverfahrens enthalten, nämlich insofern, als alle Primäreinheiten in die Stichprobe kommen, d. h. $n' = N'$ ist. Die Formel für die geschichtete Auswahl lautet

$$(42) \quad \sigma_{\mathbf{X}^2} = \sum_{i=1}^{N'} \frac{N_i^2 \sigma_i^2}{n_i} \frac{N_i - n_i}{N_i - 1}$$

Wählt man nun e_i , n_i , t_i so, daß für alle Schichten gilt

$$(43) \quad \frac{e_i^2}{t_i^2} = \frac{N_i^2 \sigma_i^2}{n_i} \frac{N_i - n_i}{N_i - 1}$$

so ist erfüllt:

$$(44) \quad \sigma_{\mathbf{X}^2} = \sum_{i=1}^{N'} \frac{N_i^2 \sigma_i^2}{n_i} \frac{N_i - n_i}{N_i - 1} = \sum_{i=1}^{N'} \frac{e_i^2}{t_i^2}$$

Mit

$$(45) \quad \frac{e^2}{t^2} = \sigma_{\mathbf{X}^2}$$

folgt also (41).

Unterstellt man zunächst, daß das innerbetriebliche Kontrollwesen in allen Teilgebieten qualitativ gleichmäßig entwickelt und wirksam ist, so daß die Sicherheitsgrade (t_i) bei der Prüfung der einzelnen Positionen ebenfalls gleichwertig anzusetzen sind, und stimmen diese Sicherheitsgrade überdies mit dem Gesamtsicherheitsgrad (t) überein, so besagt Formel (41), daß die Summe der Quadrate der Meßbeträge e_i für die Vertrauensbereiche für die Gesamtwerte der einzelnen Schichten einer Grundgesamtheit dem Quadrat des Meßbetrages e für den Gesamtvertrauensbereich höchstens gleich sein darf.

Mit dem Gesamtvertrauensbereich ist also auch der Gesamtumfang der einzelnen Vertrauensbereiche gegeben. Er ist zweckmäßig so auf die einzelnen Schichten der Grundgesamtheit zu verteilen, daß der für die Abschlußprüfung insgesamt erforderliche Stichprobenumfang minimiert wird. Das geschieht dann, wenn die e_i^2 der einzelnen Schichten

bei Vernachlässigung des Endlichkeitsfaktors in (43) proportional zum Produkt ($N_i \sigma_i$) aus Umfang und Standardabweichung der einzelnen Schichten festgesetzt werden, und

bei Berücksichtigung des Endlichkeitsfaktors in (43) so, daß die N' Verhältnisse

$$\frac{N_i^3 \sigma_i^2 (N_i - 1)}{[e_i^2 (N_i - 1) + N_i^2 \sigma_i^2 t^2]^2}$$

einander gleich werden.

Aber nicht immer sind die für die einzelnen Schichten erforderlichen Sicherheitsgrade (t_i) sowohl einander wie dem Wert (t) gleich. Wird diese Gleichheit nicht vorausgesetzt, so erhält man den *kleinsten Gesamtstichprobenumfang*, wenn

bei Vernachlässigung des Endlichkeitsfaktors die einzelnen Verhältnisse (e_i^2/t_i) proportional zum Produkt ($N_i \sigma_i$) der entsprechenden Schichten gewählt werden,

bei Berücksichtigung des Endlichkeitsfaktors die e_i so bestimmt werden, daß die Quotienten

$$\frac{t_i^2 N_i^3 \sigma_i^2 (N_i - 1)}{[e_i^2 (N_i - 1) + N_i^2 \sigma_i^2 t_i^2]^2}$$

denselben Wert annehmen.

Unter den Abschlußpositionen gibt es auch solche, für die keine Vertrauensbereiche benötigt werden, weil sie keinen Bewertungsspielraum lassen, wie z. B. das Grundkapital, die Rücklagen, die Kassen- und Bankbestände oder die wegen ihrer Bedeutung oder aus anderen Gründen, z. B. weil sie aus nur wenigen Einzelposten bestehen, lückenlos geprüft werden müssen. Besondere Vertrauensbereiche sind dann nur für die stichprobenweise zu prüfenden Abschlußpositionen zu bestimmen, also bei der Bilanz z. B. für die Halbfabrikate, Fertigfabrikate, Wertpapiere, Debitoren, Kreditoren usw., und zwar nach Maßgabe der Formel (41).

Die absoluten Umfänge der Vertrauensbereiche für die einzelnen Schichten hängen, nachdem bei gegebenen Schichtenumfängen und -varianzen der Gesamtvertrauensbereich und mit den festgesetzten Sicherheitsgraden der Schichten auch die Verhältnisse der Größen e_i für die Vertrauensbereiche der Schichten untereinander feststehen, nur noch vom *Gesamtsicherheitsgrad* (t) ab.

Bei seiner Festsetzung kann die besondere Lage des Unternehmens, der Branche und der allgemeinen Konjunkturentwicklung berücksichtigt werden, d. h. bei Unternehmen, deren wirtschaftliche Lage, insbesondere deren Rentabilität ungünstig ist, kann ein höherer und bei solchen, deren Struktur und Ertragskraft sehr günstig sind, ein geringerer Sicherheitsgrad für die Gesamtprüfung vorgegeben werden. Dadurch wird der Einfluß der Wirksamkeit des innerbetrieblichen Kontrollwesens auf die Bestimmung der besonderen Sicherheitsgrade (t_i) für die einzelnen Schichten nicht aufgehoben und auch nicht verkleinert, denn die Festsetzung eines hohen oder niedrigen Gesamtsicherheitsgrades hat zur Folge, daß der Stichprobenumfang insgesamt gehoben oder gesenkt wird.

SCHLUSS

Zusammenfassung der Ergebnisse

Die eingangs des Hauptteils gestellte Frage, ob und wo die Voraussetzungen für die Anwendbarkeit von Zufallsstichproben bei der Abschlußprüfung gegeben sind, kann aufgrund der angestellten Untersuchungen nunmehr zusammenfassend beantwortet werden.

Zunächst hat sich gezeigt, daß schärfer als bisher üblich zwischen formellen und materiellen Prüfungen unterschieden werden kann. Die Begriffe werden eindeutig, wenn man unter formellen Prüfungen solche versteht, die die Einhaltung sowohl der gesetzlichen als auch der unternehmensinternen Ordnungsvorschriften, die der Verständlichkeit und Sicherheit der Rechnungslegung dienen, zum Gegenstand haben, und wenn man als materielle Prüfungen diejenigen bezeichnet, die die Richtigkeit des Inhalts des Jahresabschlusses hinsichtlich Bewertung und Ausweis feststellen sollen. Dementsprechend können die Prüfungsfragen nunmehr in solche nach der Vollständigkeit, der formellen Ordnungsmäßigkeit (einschließlich der Einhaltung des Systems der doppelten Buchführung) und der materiellen Richtigkeit der Rechnungslegung gegliedert werden.

Die Untersuchungen, inwieweit Zufallsstichproben für die Beantwortung dieser Prüfungsfragen geeignet sind, haben zu folgenden Ergebnissen geführt:

- a) Die Prüfung der *Vollständigkeit* sowohl der Bestandsaufnahme als auch der Erfassung der Geschäftsvorfälle ist mit Hilfe von Zufallsstichproben nicht möglich. Jede Prüfung, auch die stichprobenweise, setzt die Kenntnis des Vorhandenseins der Untersuchungseinheiten voraus. Zur Abgrenzung des Prüfungsstoffes und der Verantwortlichkeit sind daher Prüfer und Statistiker gleichermaßen auf die Erteilung einer „*Vollständigkeitserklärung*“ angewiesen.
- b) Die Einhaltung des Systems der *Doppik* kann grundsätzlich mittels *Entdeckungsstichproben* geprüft werden. Mit ihrer Hilfe ist eine Aussage über die Wahrscheinlichkeit der Fehlerfreiheit der Buchführung möglich, doch ist der dazu erforderliche Auswahlatz der Stichprobe mindestens ebenso groß wie die prüferisch notwendige Wahrscheinlichkeit für die Fehlerfreiheit des Systems. Diese muß aber sehr groß sein, weil nur bei absoluter Fehlerfreiheit des Systems die durch die doppelte Verbuchung der Geschäftsvorfälle angestrebte Sicherung des Rechnungswesens erreicht wird. Entsprechend umfangreiche Stichproben sind jedoch in der Regel unrationeller als Vollerhebungen. Andererseits bieten kleine Stichproben keine ausreichende Sicherheit für die Fehlerfreiheit und können daher ebenso gut unterbleiben. Da lückenlose Prüfungen häufig aus Kostengründen nicht durchführbar sind, bleibt als Ausweg die Fehlerfeldteilung und die lückenlose Prüfung der wichtigsten Fehlerfelder, wobei ein Schluß auf die übrigen, nicht geprüften Fehlerfelder allerdings nicht möglich ist.

- c) Die Einhaltung der übrigen formellen Ordnungsvorschriften kann ebenfalls grundsätzlich stichprobenweise geprüft werden. Überschreiten die Verstöße gegen die *gesetzlichen Ordnungsvorschriften* ein vom Prüfer festgesetztes Maß, so führt dies zu prüferischen Folgerungen, wie etwa zur Unterbrechung der Prüfung, gegebenenfalls auch zur Einschränkung oder Verweigerung des Bestätigungsvermerks. Die Entscheidung darüber, ob die Buchführung in ausreichendem Maß den gesetzlichen Ordnungsvorschriften entspricht oder ob sie wegen Unterschreitung der Mindestqualität abzulehnen ist, kann mit Hilfe der *Annahmestichproben* erfolgen. Da die Verstöße gegen die gesetzlichen Ordnungsvorschriften schon einen gewissen Umfang erreichen müssen, damit sich der Prüfer veranlaßt sieht, die Buchführung als nicht ordnungsmäßig abzulehnen, brauchen die Bedingungen für die Bestimmungsfaktoren des Prüfplans keine besondere Schärfe anzunehmen, so daß in der Regel bereits mit einem verhältnismäßig kleinen Stichprobenumfang eine hinreichende Sicherheit für die prüferische Entscheidung erzielbar ist.
- d) Die Sicherheit des Rechnungswesens hängt in hohem Maße von der Güte und Zweckmäßigkeit der *unternehmensinternen Ordnungsvorschriften*, die zu einem innerbetrieblichen Kontrollsystem ausgebaut sein können, und vom Umfang der tatsächlichen Anwendung der verschiedenen Kontrollmaßnahmen ab. Während die Qualität des Systems nur durch systematische Untersuchungen, nicht aber mittels Stichproben festgestellt werden kann, ist die Wirksamkeit der einzelnen Kontrollvorschriften am Umfang der Verstöße gegen sie meßbar. Ein solches Maß ist jedoch nur gegeben, sofern Fehlerort und Fehlerart jeweils als gleichwertig anzusehen sind. Anteilsschätzungen dieser Art lassen sich rationell mittels *Schätzstichproben* durchführen, wenn die Grundgesamtheit groß genug und der Anteil fehlerhafter Einheiten in ihr nicht übermäßig groß, aber auch nicht extrem klein ist. Wo hohe Prozentzahlen fehlerhafter Einheiten anzutreffen sind, genügt bereits eine grobe Schätzung für die Entscheidung über prüferische Maßnahmen. Extrem kleine Fehleranteile erfordern sehr große Stichprobenumfänge, wenn es auf größere Aussagesicherheiten ankommt. In solchen Fällen sind Vollerhebungen oder andere organisatorische Maßnahmen vorzuziehen. Die Anwendbarkeit von Schätzstichproben erstreckt sich daher auf die Prüfung der Einhaltung derjenigen Vorschriften, die zwar in der Regel beachtet, in einem gewissen Umfang aber auch verletzt werden. Auf diesen Gebieten können solche Stichproben bei geeigneter Planung rationelle Erhebungen darstellen.
- e) Die Prüfung der Doppik und der formellen Ordnungsmäßigkeit der Rechnungslegung ermöglicht dem Prüfer ein Urteil über die Qualität der Buchführung. Diese kann er an der Zahl der in ihr enthaltenen fehlerhaften Einheiten objektiv messen. Aufgrund des Grades der Ordnungsmäßigkeit und Zuverlässigkeit kann der Prüfer dann über die *Festsetzung des Sicherheitsgrades (t)* in den bei den materiellen Prüfungen zu verwendenden Formeln für Schätzstichproben den Umfang der materiellen Prüfungen beeinflussen. Allerdings können die verschiedenen Fehlerarten und Fehlerorte von unterschiedlicher Bedeutung für die Sicherheit der Rechnungslegung sein. Mit Rücksicht hierauf ist

eine Gesamtbewertung der Sicherheit der Buchführung nicht ohne weiteres möglich. Sie ist aber auch nicht erforderlich, da sich die materiellen Prüfungen jeweils auf die einzelnen Teilprüfungsgebiete, in der Regel auf die einzelnen Abschlußpositionen, erstrecken. Infolgedessen genügt eine Bewertung der Wirksamkeit der innerbetrieblichen Kontrollmaßnahmen innerhalb des jeweiligen Teilgebietes für die Bestimmung des Ausmaßes der materiellen Prüfungen. Die Feststellung kleiner Fehleranteile in solchen Teilgebieten bewirkt ein höheres Vertrauen in die Richtigkeit dieser Aufzeichnungen und läßt daher eine Einschränkung der materiellen Prüfungen über eine entsprechende Herabsetzung des Sicherheitsgrades zu, während umgekehrt größere Fehleranteile zu einer entsprechend intensiveren materiellen Prüfung der betreffenden Teilgebiete führen müssen.

- f) Die bei der Prüfung der *materiellen Richtigkeit* der Rechnungslegung auftretenden Prüfungsfragen können grundsätzlich alle mittels *Schätzstichproben* gelöst werden. Bei der Prüfung der *Bestände* interessieren in erster Linie die *Gesamtwerte* der einzelnen Bilanzpositionen. Soweit *Analysen von Beständen* in Frage kommen, interessieren *Mittelwerte* oder *Anteile* von Merkmalen oder Merkmalkategorien. Alle diese benötigten Werte braucht der Prüfer nicht mit letzter Genauigkeit zu kennen. Unterhalb des gesetzlichen Höchstwertes besteht für die Bewertung der Vermögensposten ein erheblicher Ermessensspielraum, während Überschreitungen des gesetzlichen Höchstwerts mit einer solchen Genauigkeit ermittelt werden müssen, daß Herabsetzungen bis mindestens auf die gesetzlichen Höchstwerte erfolgen können. Die Genauigkeit der Schätzungen von Mittelwerten und Anteilen ist so zu bestimmen, daß die Zwecke, die die Analyse im Rahmen der Prüfung zu erfüllen hat, erreicht werden. Durch eine den Verhältnissen angepaßte Festsetzung des Sicherheitsgrades und des Vertrauensbereichs sowie des Auswahlverfahrens können die prüferischen Erfordernisse mit rationellen Stichprobenumfängen erfüllt werden. Die relativen Stichprobenumfänge sind desto kleiner, je größer die Umfänge der Grundgesamtheiten und je kleiner die Varianzen in ihnen sind.
- g) Auch die bei der Prüfung der *Geschäftsvorfälle* auftretenden Fragen lassen sich mit Hilfe von *Schätzstichproben* beantworten, da es sich ebenfalls um die Ermittlung von Gesamtwerten, Mittelwerten und Anteilen handelt. Auf diesem Prüfungsgebiet begegnet man häufig großen Buchungsmassen mit ähnlichem Inhalt, d. h. mit kleinen Varianzen unter den Merkmalwerten. Hier sind dann Schätzstichproben besonders geeignete Prüfungsverfahren. Erfolgt die Verbuchung solcher Geschäftsvorfälle mittels selbsttätig arbeitender Geräte, so kann man von einer relativ kleinen Anzahl von Belegen und den ihnen zugrunde liegenden Geschäftsvorfällen, unter Überspringen der Prüfungen der Übernahme der Belegdaten in die Maschine, der Programmierung und der Arbeitsweise der Maschine, unmittelbar auf das Ergebnis der Rechnungslegung über die gesamten gleichartigen Geschäftsvorfälle, wenn auch nicht genau, so doch mit einem vom Prüfer nach fachlichem Ermessen bestimmbaren Sicherheitsgrad und Vertrauensbereich schließen. Damit bietet das Zufallsstichprobenverfahren einen von den Geräten, ihren Programmen sowie den Kontrolleinrichtungen und -aufzeichnungen unabhängigen Prüfungsweg.

- h) Schließlich ergaben die Untersuchungen, daß für die Festsetzung des für die Stichproben benötigten „*Vertrauensbereichs der Abschlußprüfung*“ die Höhe des *Jahreserfolgs* oder die Höhe des *Eigenkapitals* oder Teilgrößen hiervon als objektive Bezugsgrößen herangezogen werden können. Eine Verteilung dieses Gesamtvertrauensbereichs auf die einzelnen Bilanzpositionen bzw. auf die einzelnen Positionen der Gewinn- und Verlustrechnung ist möglich, wenn man die in den einzelnen Bilanzpositionen zusammengefaßten Bestandsposten als Teilgesamtheiten der Grundgesamtheit „Bestände“ und die in den einzelnen Positionen der Gewinn- und Verlustrechnung als Teilgesamtheiten der Grundgesamtheit „erfolgswirksame Geschäftsvorfälle“ ansieht und aus beiden Grundgesamtheiten geschichtete Stichproben zieht. Dann kann ein „*Sicherheitsgrad der Abschlußprüfung*“ vorgegeben werden, dessen Höhe sich nach der besonderen Lage des Unternehmens, der Branche und der allgemeinen Konjunktur-entwicklung richten kann, allerdings ohne daß sich hierfür objektive Maßstäbe angeben lassen.

Literaturverzeichnis

- Adler, H./Düring, W./Schmaltz, K.* Rechnungslegung und Prüfung der Aktiengesellschaft, 3. Aufl., Stuttgart 1957
- American Institute of Certified Public Accountants* Generally Accepted Auditing Standards, Special Report by the Committee on Auditing Procedure, 1954
- Internal Control, Special Report by the Committee on Auditing Procedure, 1949
- Glossary of Statistical Terms for Accountants and Bibliography on the Application of Statistical Methods to Accounting, Auditing and Management Control, 8. September 1958
- Anderson, O. N.* Probleme der statistischen Methodenlehre in den Sozialwissenschaften, 3. Aufl., Würzburg 1957
- Arkin, H.* Statistical Sampling in Auditing, New York Certified Public Accountant, Juli 1957, S. 454 ff.
- A Statistician Looks at Accounting, Journal of Accountancy, April 1958, S. 66 ff.
- Discovery Sampling in Auditing, Journal of Accountancy, Februar 1961, S. 51 ff.
- Auffermann, J. D.* Jahresabschlußprüfung, Handwörterbuch der Betriebswirtschaft, 3. Aufl., Stuttgart 1958, Bd. 2, Spalte 2896 ff.
- Backer, M./Fertig, P. E.* Statistical Sampling and the Accounting Curriculum, The Accounting Review 1958, S. 415 ff.
- Baudisch, R.* Die Funktion des Abschlußprüfers, Wirtschaftsprüfung als Element gesunder Wirtschaftsentwicklung, Düsseldorf 1955, S. 19 ff.
- Berg, K.* Wie wird geprüft?, 6. Aufl., Bonn 1949
- Boitte, A.* Statistique et revision, Journal de la Comptabilité (belg.), 20. Jg., 1. Oktober 1958, 21. Jg., Februar 1959
- Les règles standard de revision et leur application, Journal de la Comptabilité, 21. Jg., März 1959, und 21. Jg., Mai 1959
- Brink, V. Z.* The Independent Auditor's Review of Internal Control, Journal of Accountancy 1942, Bd. 73, S. 430 ff.
- Statistical Sampling, Internal Auditing, 1958, S. 339 ff.

- Brown, R. G.* Statistical Sampling Tables for Auditors, Journal of Accountancy, Mai 1961, S. 46 ff.
- Brown, R. G./Vance, L. L.* Sampling Tables for Estimating Error Rates and Other Proportions, Berkely 1961
- Buchner, R.* Bewußte Auswahl oder Zufallsauswahl bei der aktienrechtlichen Pflichtprüfung. Die Wirtschaftsprüfung, Dezember 1961, Jg. 14, Nr. 24, S. 657 ff.
- Bujack, H./Roth, L.* Schätzverfahren in der Inventur, Der Betrieb 1959, Jg. 12, Nr. 21, S. 577 ff., und Nr. 22, S. 601 ff. Mathematisch-statistische Methoden bei der Inventur der Werkstattbestände, hrsg. v. d. Deutschen Arbeitsgemeinschaft für statistische Qualitätskontrolle beim AWF, Frankfurt 1962
- Cameron, J. M.* Tables for Constructing and for Computing the Operation Characteristics of Single Sampling Plans, Industrial Quality Control, Juli 1952, S. 37 ff.
- Carman, L. A.* The Efficacy of Tests, The American Accountant, Jg. 1933, S. 360 ff.
- Cochran, W. G.* Sampling Techniques, New York—London 1953
- Coulinet, J.* Application de la technique des sondages à la revision comptable, Rapport de synthèse der Verhandlungen der Commission Mathématiques Financières der UEC, Nizza 1958 (unveröffentlichtes Manuskript)
Applications de la statistique à la gestion des entreprises, Aufzeichnung — Zürich 1961 (unveröffentlichtes Manuskript)
- Courtright, W. E./Procassini, A. A.* Inventory Valuation by Sampling, Industrial Quality Control, 1958, Bd. 4, Nr. 8, S. 16 ff.
- Dalenius, T.* Technik und Methode der Stichprobenerhebungen, hrsg. v. Fachausschuß Stichprobenverfahren der Deutschen Statistischen Gesellschaft, München 1950
- Davis, D. D./Rounsaville, A.* Auditing Voluminous Data by Modern Sampling Methods, Journal of Accountancy, Juni 1959, S. 45 ff.
- Deming, W. E.* Some Theory of Sampling, New York — London 1950
On Simplifications of Sampling Design through Replication With Equal Probabilities and Without Stages, Journal of the American Statistical Association, 1956, Bd. 51, S. 24 ff.
Sample Design in Business Research, New York — London 1960

- Dinter, H.* Zur Frage der Verwendung von Stichprobenverfahren im Rahmen von Prüfungshandlungen, Die Wirtschaftsprüfung, Februar 1962, Jg. 15, Nr. 4, S. 85 ff.
- Dodge, H. F./Romig, H. G.* Sampling Inspection Tables, 3. Aufl., New York — London 1946
- Dunn, K. W.* Clarification of Responsibility for Discovery of Fraud, Journal of Accountancy, August 1958, S. 49 ff.
- Farnsworth, W. P./Teitelbaum, L. N.* Can Auditors Use Statistical Sampling Methods? The Federal Accountant, Juni 1957, S. 17 ff.
- Fischer, B.* Die formelle Prüfung, Die Wirtschaftsprüfung, 1952, Jg. 5, S. 121 ff.
- Guinn, W. E.* Error Sampling in Auditing, National Accountants Association, Bulletin, 1958, Bd. 39, Nr. 5, S. 12
- Haas, G./Oechsner, L.* Der Jahresabschluß nach Handels- und Steuerrecht, Wiesbaden 1958
- Hagest, K.* Grundsätze ordnungsmäßiger Abschlußprüfung, Die Wirtschaftsprüfung, 1949, Jg. 2, S. 450 ff.
- Hansen, M. H./Hurwitz, W. N./Madow, W. G.* Sample Survey Methods and Theory, 3. Aufl., New York — London 1960
- Heine, K. H.* Der Prüfungsplan für die Prüfung der Ordnungsmäßigkeit der Buchführung, Die Wirtschaftsprüfung, 1962, Jg. 15, Nr. 2, S. 29 ff.
- Hertlein, A./Meisner, K.* Abschluß und Prüfung der Unternehmungen einschließlich Steuerprüfung, 4. Aufl., Wiesbaden 1956
- Hill, H. P.* An Accountant Looks at Statistics, Journal of Accountancy, April 1958, S. 57 ff.
- Hintner, O.* Praxis der Wirtschaftsprüfung, 3. Aufl., Stuttgart 1949
- Institut der Wirtschaftsprüfer (Hrsg.)* Die Fachgutachten und Stellungnahmen des Instituts der Wirtschaftsprüfer auf dem Gebiete der Rechnungslegung und Prüfung, Düsseldorf 1956
Wirtschaftsprüfer-Handbuch 1959, Düsseldorf 1959
Mindestanforderungen zur Sicherung der Ordnungsmäßigkeit kaufmännischer Rechnungslegung beim Einsatz selbsttätig arbeitender Geräte, Die Wirtschaftsprüfung, 1961, Jg. 14, Nr. 11, S. 308
- Isaac, A.* Revision und Kontrolle, Wiesbaden 1951
- Jones, H. L.* Some Problems in Sampling Accounting Records, Quality Control Conference Papers, 1952, S. 209 ff.

- Johnson, R. W.* The Use and Significance of Random Samples in Audit Tests, *Journal of Accountancy*, Dezember 1957, S. 43 ff.
- Kellerer, H.* Theorie und Technik des Stichprobenverfahrens, Einzelschriften der Deutschen Statistischen Gesellschaft, Nr. 5, 2. Aufl., München 1953
Stichprobenverfahren, Handwörterbuch der Sozialwissenschaften, Stuttgart, Tübingen, Göttingen 1959, Bd. 10, S. 193 ff.
Statistik im modernen Wirtschafts- und Sozialleben, rowohlt's deutsche enzyklopädie, Reinbek bei Hamburg 1960
- Kendall, M. G./
Buckland, W. R.
Klein, A.* A Dictionary of Statistical Terms, 2. Aufl., Edinburgh — London 1960
Über die Wahrscheinlichkeit der Entdeckung von Fehlern bei Revisionen, *Zeitschrift für handelswissenschaftliche Forschung*, Jg. 1911/12, S. 580 ff.
Gesetz über Aktiengesellschaften und Kommanditgesellschaften auf Aktien (Aktiengesetz) vom 30. 1. 1937 (RGBl. I, S. 107) und Einführungsgesetz zum Gesetz über AGs und KGaA vom 30. 1. 1937 (RGBl. I, S. 166) nebst vollständiger Begründung, Stuttgart — Berlin 1937
- Küch, O.* Treuhand- und Revisionswesen, Stuttgart 1961
Statistical Sampling in Accountancy, *The Accountant*, Oktober 1958, S. 491 ff. und November 1958, S. 524 ff.
- Loitlsberger, E.
Mackenzie, H. C.* Die Durchführung der Revision, Probleme neuzeitlicher Unternehmensführung, Arbeitsunterlagen für den IV. Kongreß der UEC vom 18. bis 21. 9. 1961 in Zürich, S. 137 ff.
- Meier, A.* Rechtliche und tatsächliche Gestaltung der Jahresabschlußprüfung von Aktiengesellschaften, *Zeitschrift für handelswissenschaftliche Forschung*, Neue Folge, Jg. 9, 1957, S. 376 ff.
- Minz, G.* Die formalen Ordnungsprinzipien der Buchführung, Diss., Köln 1956
- Minz, W.* Prüfungsmethoden, *Die Wirtschaftsprüfung*, 1960, Jg. 13, Nr. 4, S. 89 ff.
- Nertinger, J.* Wie prüft man einen Jahresabschluß?, Stuttgart 1949
- Obrock, R. F.* A Case Study of Statistical Sampling, *Journal of Accountancy*, März 1958, S. 53 ff.
- Peterson, O. S.* Statistisk stikproveteknik anvendt i revisionsarbejdet, Meddelelser fra foreningerne af statsautoriserede revisorer, hrsg. vom Foreningens Hovedbestyrelse, Nr. 1 vom 15. 1. 1960, S. 5 ff.

- Peupelmann, H. W.* Deutsche oder amerikanische Prüfungsmethoden, Die Wirtschaftsprüfung, 1959, Jg. 12, Nr. 14, S. 382 ff.
- Pinner, A.* Aktiennovelle und Bankenaufsicht — Verordnung vom 19. 9. 1931, Berlin — Leipzig 1931
- Pougin, E.* Die Berücksichtigung des internen Kontrollsystems als Grundlage ordnungsmäßiger Abschlußprüfung, Düsseldorf 1959
Zur gesetzlichen Neuregelung des Umfangs der aktienrechtlichen Jahresabschlußprüfung, Die Wirtschaftsprüfung, 1959, Jg. 12, Nr. 13, S. 350 ff.
Grundsätze ordnungsmäßiger Abschlußprüfung in USA und Deutschland, Die Wirtschaftsprüfung, 1960, Jg. 13, Nr. 1, S. 2 ff.
- Rackles, R.* Das Problem der repräsentativen Auswahl bei der aktienrechtlichen Abschlußprüfung, Frankfurt/M. 1961
- Reichsgericht* Entscheidungen in Zivilsachen, Bd. 116, S. 119, Urteil vom 15. 10. 1937 (II 61/37)
- Reiniger, W.* Zur Reform der Vorschriften über Publizität, Rechnungslegung und Prüfung der Aktiengesellschaften, Diss., Düsseldorf 1957
- Ridilla, R. A.* A Simplified Statistical Technique for Use in Verifying Accounts Receivable, The Accounting Review, Oktober 1959, S. 547 ff.
- Rowles, B. M.* Statistical Techniques in Accounting, Accounting Papers of the Eleventh Annual Conference of Accountants, University of Fulsa 1957
- Rudell, A. L.* Applied Sampling Doubles Inventory Accuracy Halves Cost, National Accountants Association, Bulletin, Oktober 1957, S. 5 ff.
- Saage, G.* Die stillen Reserven im Rahmen der aktienrechtlichen Pflichtprüfung, Köln-Opladen 1959
- Schaafsma, A. H./*
Willemze, E. G. Moderne Qualitätskontrolle, 2. Aufl., Eindhoven 1959
- Schlegelberger, F./*
Quassowski, L., u. a. Aktiengesetz, 3. Aufl., Berlin 1939
- Schmalenbach, E.* Über Einrichtungen gegen Unterschlagungen und über Unterschlagungsrevision, Zeitschrift für handelswissenschaftliche Forschung, Jg. 1911/12, S. 321 ff.
- Schmölder, K.* Rechnungslegung, Prüfung und Publizität der Aktiengesellschaften, Zeitschrift für handelswissenschaftliche Forschung 1950, Neue Folge, Jg. 2, S. 279 ff.

- Schröder, J.* Die Auswirkungen des Einsatzes von datenverarbeitenden Maschinen in den Unternehmungen auf die steuerliche Betriebsprüfung, Die steuerliche Betriebsprüfung 1961, Jg. 1, Heft 1, S. 3 ff.
- Shonting, D. M./Stone, L. D.* Audit Techniques for Electronic Systems, Journal of Accountancy, Oktober 1958, S. 54 ff.
- Sprague, W. D.* Testing the Adequacy of Internal Control, Journal of Accountancy, 1956, Bd. 101, S. 49 ff.
- Statistisches Bundesamt (Hrsg.)* Stichproben in der amtlichen Statistik, Stuttgart — Mainz 1960
- Strauß, W.* Das Aktienrecht in der sozialen Marktwirtschaft, Frankfurter Allgemeine Zeitung, Nr. 26 vom 31. 1. 1959, S. 5
- Stringer, K. W.* Some Basic Concepts of Statistical Sampling in Auditing, Journal of Accountancy, November 1961, S. 63 ff.
- Szameitat, K./Koller, S.* Über den Umfang und die Genauigkeit von Stichproben, Wirtschaft und Statistik, 1958, Neue Folge, Jg. 10, S. 10 ff.
- Trueblood, R. M./Cyert, R. M.* Sampling Techniques in Accounting, New York 1957
- Trumpler, H.* Die Bilanz der Aktiengesellschaft, Basel 1950
- Union Européenne des Experts Comptables Economiques et Financiers* Unternehmen und Unternehmensberater im Gemeinsamen Europäischen Markt, Bericht über den Dritten Kongreß der UEC in Nizza 1958, Düsseldorf 1960
- Probleme neuzeitlicher Unternehmensführung, Arbeitsunterlagen für den Vierten Kongreß der UEC vom 18. bis 21. 9. 1961 in Zürich, Die Prüfung des Jahresabschlusses, Düsseldorf 1961
- Lexique UEC Lexicon, Wörterbuch des Rechnungswesens, Düsseldorf 1961
- Vance, L. L./Neter, J.* Statistical Sampling for Auditors and Accountants, New York — London 1956
- Voigt, A.* Prüfungstechnik und Prüfungsbericht, Aktuelle Bilanz- und Steuerprobleme, Beiheft 3 zu Die Wirtschaftsprüfung, 1949, S. 130 ff.
- Wald, A.* Sequential Analysis, New York 1947
- Wipper, R.* Revisionstechnik und Buchführungsverfahren, 3. Aufl., Stuttgart 1951
- Yates, F.* Sampling Methods for Censuses and Surveys, 3. Aufl., London 1960
- Zimmerman, E.* Theorie und Praxis der Prüfungen im Betrieb, Essen 1954

Lebenslauf

Ich wurde am 9. März 1932 in Berlin geboren. Nach Besuch der Volksschule und höherer Lehranstalten in Berlin, Belzig, Recklinghausen und Mülheim-Ruhr erhielt ich Ostern 1952 das Reifezeugnis des naturwissenschaftlichen Gymnasiums in Mülheim-Ruhr. In den Jahren 1952 bis 1954 war ich kaufmännischer Volontär bei Unternehmen verschiedener Wirtschaftszweige (Wirtschaftsprüfungsgesellschaft, Edelstahlwerk, Eisengroßhandlung, Kreditbank). Ab Wintersemester 1952/53 studierte ich Wirtschaftswissenschaften an den Universitäten Köln, Freiburg und München, wo ich im Februar 1958 die Diplomprüfung für Kaufleute bestand. Anschließend war ich als Prüfungsassistent bei einer Treuhandstelle für Wohnungsunternehmen tätig. Aufgrund meines Interesses an prüferischen und statistischen Fragen befaßte ich mich seit 1959 mit dem Thema meiner Dissertation. Im Jahre 1962 erweiterte ich meine praktischen Kenntnisse in einer Wirtschaftsprüfungsgesellschaft.

